The sasquatch Mouse: an Enhanced Limb

Simon J.H. Heaney

PhD – The University of Edinburgh – 2002

Declaration

I declare that,

a) this thesis is composed by myself.

b) this work is my own, except where otherwise stated.

Simon Heaney

October 2002

Table of Contents

List of Figures

List of Tables

List of Abbreviations

°C	degrees centigrade
µg	micrograms
µl	microlitres
µM	micromolar
A-P	anterior to posterior
AER	apical ectodermal ridge
BAC	bacterial artificial chromosome
bp	base-pairs
BMP	bone morphogenetic protein
BSA	bovine serum albumin
Ca^{2+}	calcium ion
cDNA	complementary deoxyribonucleic acid
cM	centiMorgan
CNS	conserved non-coding sequence
dH_2O	distilled water
D-V	dorsal to ventral
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
E	embryonic day
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetic acid
EST	expressed sequence tag
F1	first generation offspring
FGF	fibroblast growth factor
g	grams
g	relative centrifugal force
HGMP	Human Genome Mapping Project
HPAP	*human placental alkaline phosphatase gene*
HPE	holoprosencephaly
IM	Intermediate mesoderm,
K^+	potassium ion
kb	kilobase-pairs
l	litres
LacZ	β-galactosidase gene
LPM	lateral plate mesoderm
M	molar
Mb	megabase-pairs
mg	milligrams
ml	millilitres
mM	millimolar
mRNA	messanger ribonucleic acid
nt	nucleotides
P-D	proximal to distal
PAC	P1 artificial chromosome
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PPD	pre-axial polydactyly

q	human chromosome long arm
RNA	ribonucleic acid
RT-PCR	reverse-transcriptase polymerase chain reaction
SDS	sodium dodecyl (lauryl) sulphate
Shh	*sonic hedgehog*
Ssq	sasquatch
SSC	saline sodium citrate
Tris	tris hydroxymethyl aminomethane
UTR	untranslated region
ZPA	zone of polarising activity

Acknowledgements

Firstly I would like to thank my supervisor Robert (Bob) Hill. I am grateful for his guidance, encouragement and patience throughout my PhD, and I hope his American flag model of A-P limb patterning gains the recognition it deserves. Laura Lettice provided help with all things molecular, Lorna Purdie guided me through the intricacies of transgenic work and the other members of the lab, Rob Watson, Alison Wright, and Carlo Deangelis, have all provided invaluable assistance.

The staff of both the transgenic and BRF mouse houses provided essential help with the mice used in this project, and I would like thank Lisa Mckie, William Mungle and Brendan Doe in particular. The high degree of professionalism and care with which my mice were handled in both mouse houses made this PhD possible. Stewart Mckay and Agnes Gallacher provided excellent technical assistance with the *Fugu* sequencing.

I would also like to thank James Sharpe, the original creator of the sasquatch mouse, for the stimulating discussions we had and help with the *HPAP* staining. Martin Taylor and Colin Semple provided invaluable guidance with the Bioinformatics contained within this thesis, and their capacity to consume alcohol was awe-inspiring.

Fellow members of C3 who kept me sane during my project include Mark Rolfe, Lee Spraggon and Duncan Whitaker who are respectively, the most offensive, crude and unluckiest people I know. Heather MacPherson and Shirley Smith proof read large portions of this thesis, providing encouragement and insightful comments. The rest of C3 also deserve my thanks for the consideration they showed me during my PhD. Sandy Bruce, provided invaluable assistance with many of the figures, and Dawn Loh helped to format and edit this thesis when all seemed lost.

Collaborators who generously shared their data with me to make the story contained within this body of work more interesting include Debbie Goode, Greg Elgar and Esther de Graaff.

Finally I would like to thank my Mum, Dad and sisters for their support, both financial and emotional during my PhD. This thesis is dedicated to my Dad, who could not see its completion, but who never doubted my ability.

Abstract

The sasquatch (*Ssq*) mouse is a member of the hemimelia-luxate group of mutations, which display a variety of limb defects including preaxial polydactyly (extra digits). *Ssq* was generated serendipitously via a random transgenic insertion of a reporter construct approx. 1Mb downstream of the gene sonic hedgehog *(Shh).* Associated with the limb defects in all members of the hemimelia-luxate group of mutations is the mis-regulation of *(Shh) i*n the limb bud, resulting in an anterior area of ectopic expression as well as the normal patch of posterior expression. This process of *Shh* mis-regulation is poorly understood for most of the hemimelia-luxate mutants, but recently we have determined that in the case of the *Ssq* mouse, disruption of a long range *Shh* limb enhancer is responsible for ectopic *Shh* expression.

The presence of an enhancer was initially revealed by the *Ssq* transgenic reporter gene, which displayed an expression pattern virtually identical to *Shh* in both normal and affected limb buds. Subsequently we isolated the transgenic insertion to an intron of the gene *Lmbr1*. To ascertain a direct or indirect role for this enhancer we devised and implemented a genetic assay the "*cis-trans* test". Using meiotic recombination events we generated several mice which recombined a *Shh* null locus to a *Ssq* locus on the same chromosome (i.e. in *cis*). These "recombinant mice" demonstrated complete suppression of the *Ssq* phenotype whereas mice carrying *Ssq* and *Shh* loci on opposite homologous chromosomes (i.e. in *trans*) showed a *Ssq* heterozygous phenotype. Thus the *Ssq* locus can only stimulate ectopic *Shh* expression from the *Shh* locus located on the same chromosome as itself (i.e. in *cis*); leading to the conclusion that the Ssq mutation has disrupted the activity of a long-range *Shh* limb enhancer rather than a gene.

To identify the *cis*-acting elements implicated in the *Ssq* mutation a comparative sequence analysis of the *Lmbr1* genomic region in human, mouse and *Fugu* was implemented. Using the Vista and Pipmaker software packages, several highly conserved non-coding genomic sequences (CNSs) were identified as candidate regulatory elements. Two CNS regions that lie close to the *Ssq* insertion site were cloned into *LacZ* reporter constructs designed to assay for enhancer activity. Injection of these constructs into single cell mouse embryos to generate transgenic mice revealed that one of the CNS regions is capable of driving *LacZ* in an expression pattern identical to that of *Shh* in the developing limb bud. We believe that this CNS is an enhancer element responsible for driving *Shh* expression in the developing limb bud, and disruption of this element by the *Ssq* insertion is responsible for the phenotype observed in *Ssq* mice.

The human genetic disease Preaxial Polydactyly (PPD) maps to a 450kb region syntenic to the *Ssq* insertion site. PPD is also a limb specific defect resulting in extra digits and no coding or splicing mutations have been found within the PPD critical region. *Ssq* is most likely the model for PPD confirming the likelihood that enhancer elements responsible for human genetic disease or variation could be acting over 100's of kb within the human genome. Obviously this has important consequences for the mapping of human genetic disease and reaffirms the importance of mouse models in revealing underlying disease mechanisms.

Chapter 1

Introduction

1.1 The vertebrate limb

The development of paired limbs at appropriate levels along the primary body axis is a hallmark of the body plan of jawed vertebrates. Cells from the lateral plate mesoderm (LPM) that contribute to the developing limb are specified early in development, before the manifestation of any physical limb like structures and are termed the limb field (Johnson and Tabin, 1997) (see Fig1.1). A clue as to how LPM cells are recruited into the limb fields came from foil barrier and extirpation studies; the removal or separation of underlying intermediate mesoderm (IM) from the LPM inhibited limb formation (Geduspan and Solursh, 1992). The conventional interpretation of these experiments is that the IM produces a factor that induces and maintains the limb field in the LPM. It should be noted that a more recent study examining the role of the IM in limb bud initiation has produced conflicting results (Fernandez-Teran et al., 1997). However, what is clear is that once specified, cells in the limb field contain all the information necessary to generate a limb. This is neatly demonstrated by the transplantation of limb field tissue into areas of the flank not normally associated with limb development, which results in the generation of complete ectopic limbs

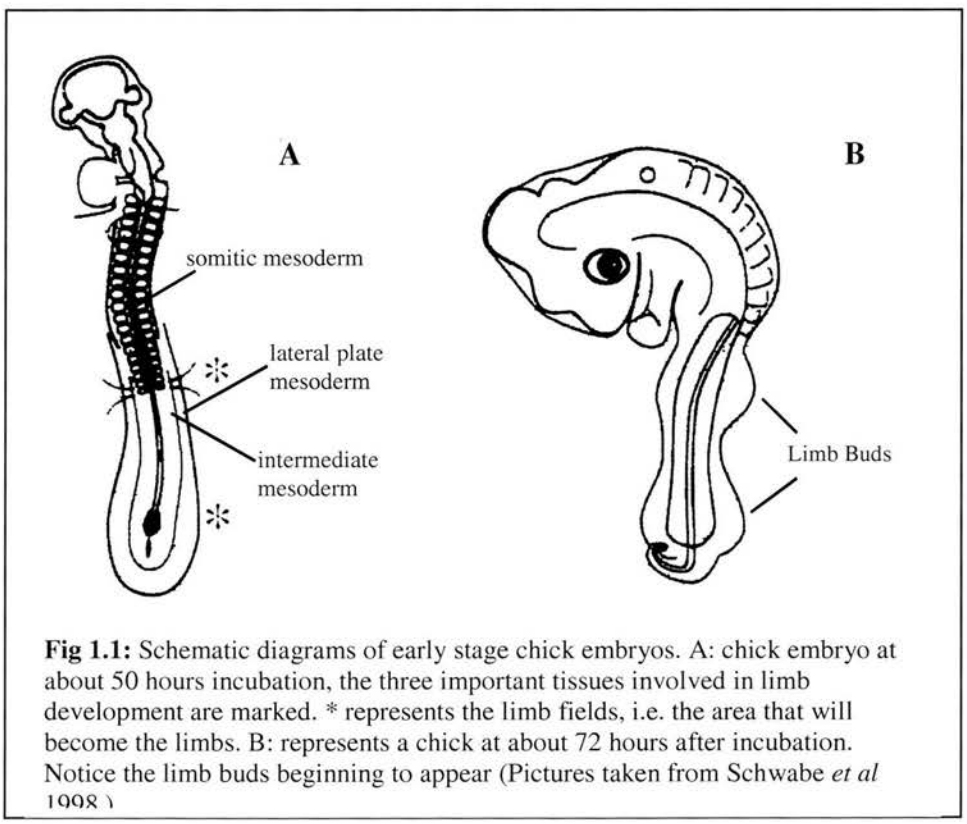


Fig 1.1: Schematic diagrams of early stage chick embryos. A: chick embryo at about 50 hours incubation, the three important tissues involved in limb development are marked. * represents the limb fields, i.e. the area that will become the limbs. B: represents a chick at about 72 hours after incubation. Notice the limb buds beginning to appear (Pictures taken from Schwabe *et al* 1998)

(Schwabe et al., 1998).

The first physical signs of the developing limb are protrusions from the lateral body wall of the embryo known as limb buds (see Fig 1.1), which are composed of cells from the LPM and somites (Cohn and Bright, 1999). LPM cells proliferate and eventually give rise to the skeletal elements and connective tissue of the limb. Cells from the lateral edges of nearby somites migrate into the limb bud shortly after its initiation, ultimately differentiating to form the muscles and vasculature of the limb(Johnson and Tabin, 1997).

As development progresses the limb bud exhibits morphological changes across 3 axis. Outgrowth of the limb occurs across the proximal-distal axis of the limb (P-D), shoulder to fingers, the limb becomes flattened across the dorsal-ventral axis (D-V), palm to back of the hand, and asymmetric structures emerge across the anterior to posterior axis A-P, thumb to little finger. Limb structures become apparent with the passage of time, the most proximal elements, stylopod, differentiate first, followed by the progressive differentiation of ever more distal structures that make up the zeugopod and eventually the autopod (Johnson and Tabin, 1997) (see Fig 1.2).

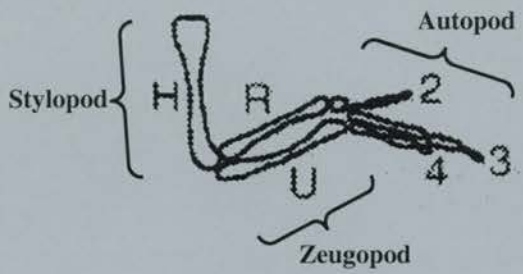


Fig 1.2: Chick wing showing the major P-D skeletal structures. Stylopod consists of the humerus, the Zeugopod the radius and ulna, and finally the autopod contains the digits and wrist bones. (Limb drawing taken from Schwabe *et al.*1998)

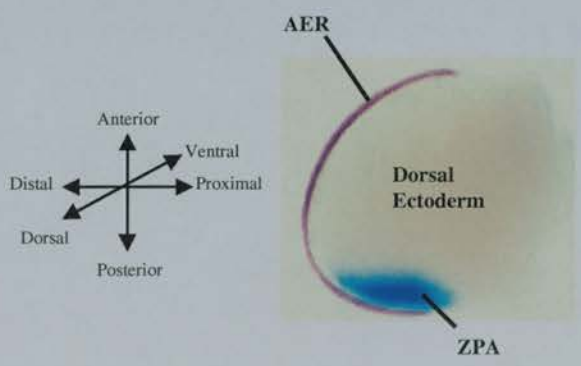


Fig 1.3: The three major limb signalling regions, depicted on a E10.5 mouse limb bud. The view is looking down onto the dorsal ectoderm.

The process by which cells that make up the limb establish their position relative to these three axis and therefore their ultimate fate is termed pattern

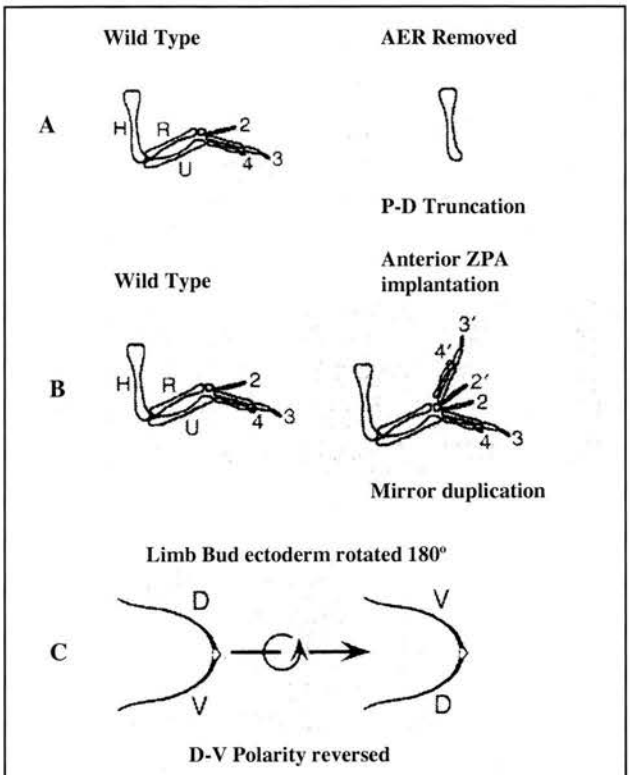


Fig 1.4: The surgical manipulations that led to the discovery of the major limb signalling regions. **A** depicts the removal of the AER, which results in truncations suggesting that the AER provides signals essential for outgrowth. **B** mirror image duplications generate by grafting ZPA tissue, suggesting the ZPA has a role in A-P patterning. **C** diagrammatic representation of the reversal of the reversal in D-V polarity created upon the rotation of the limb ectoderm by 180°, suggesting that the ectoderm provides D-V polarity information. (Taken from Schwabe *et al.*1998)

formation, and is believed to rely on three distinct signalling centres within the limb bud(Johnson and Tabin, 1997) (see Fig 1.3). The first of which is the apical ectodermal ridge (AER), a morphologically distinct epithelial structure that runs from the anterior to the posterior of the distal-most part of the limb bud. Surgical removal of the AER from developing chick limb buds results in a failure of the limb to grow along the P-D axis (Saunders, 1948) (see Fig 1.4). The earlier the removal occurs during development the greater the loss of distal structures, indicating that signals from the AER are crucial to maintain outgrowth in the limb bud. However, the remaining structures in limbs truncated by the loss of an AER still exhibit A-P and D-V polarity, suggesting that information with regards to these axis is derived from other sources in the limb.

A second, morphologically indistinct, signalling region of the limb bud was identified by the grafting of small areas of posterior limb bud mesenchyme into the anterior region of limb buds (Saunders, 1968). These grafts resulted in mirror image duplications of the distal limb, generating posterior structures in the anterior of the limb bud (see Fig 1.4). Because of the graft's ability to re-organise the A-P polarity of the limb, this posterior region

of the limb was termed the zone of polarising activity (ZPA) and was believed to be the source of signals that determined cell fates along the A-P axis of the limb.

A third set of surgical experiments (Geduspan and MacCabe, 1987) rotated the limb-bud ectoderm 180°, resulting in a reversal of the D-V polarity of distal limb structures, indicating that signals from the limb bud ectoderm were responsible for regulating the D-V patterning of the limb (see Fig 1.4). Thus it was believed that molecular cues from the AER drove the P-D outgrowth of limb tissue, which was provided with positional information relative to the A-P and D-V axis of the limb by molecular signals from the ZPA and limb bud ectoderm. The identification of these limb signalling regions was a major leap forward in understanding the processes of development, but the identification of the molecules responsible for their activity has only recently been possible with the advent of modern genetic and molecular biology techniques.

1.2 Specification of the Limb Field

The axial level at which the limb field is specified is obviously extremely important for the embryo to get right; arms and legs must develop from appropriate places along the flank. The *HOX* genes generate the major source of A-P positional information along the main body axis of the embryo (Favier and Dolle, 1997), and several lines of evidence point to them having a role in positioning the limb field.

Firstly their nested patterns of expression along the A-P axis (the *HOX* code) is thought to control the identity of axial structures (Favier and Dolle, 1997). The *HOX9* paralogues in particular have an expression pattern consistent with a role in specifying the axial position of the forelimb (Cohn et al., 1997). Secondly in species with different axial morphologies, such as snakes, changes in *HOX* expression correlate with morphological changes in body plan. Studies in python embryos (Cohn and Tickle, 1999) have demonstrated that the *HOXC6* and *HOXC8* genes, which promote thoracic identity, expand their expression domains to encompass most of the axial skeleton. Thus the limb-less python

embryo does not display the *HOX* coding normally associated with limb development. Thirdly the experimental generation of ectopic limbs, results in a repatterning of the *HOX* code to mimic that of the endogenous limbs (Cohn et al., 1997). Finally a loss of function mutation of *HOXB5* can shift the axial position of the developing limb (Rancourt et al., 1995).

Although it is likely that the *HOX* genes provide the initial positional information to place the limb fields, other factors must interpret this information to actually begin the process of limb development. Several genes have been identified as possible initiators of limb development based on their early expression in the limb field (Isaac et al., 2000). *SnR,* a zinc finger transcription factor and orthologue of the *Drosophila* gene *Snail,* is believed to play an important role in mesoderm formation (Grau Y, 1984). Together with another transcription factor *twist, SnR* is expressed in the limb fields of both chick wings and legs early in development (Sefton et al., 1998). Both genes have been suggested to play a role in the formation of *Drosophilia* wing discs (Fuse et al., 1996; Emori and Saigo, 1993), and outgrowth of chick limbs (Kanegae et al., 1998), however a precise function has yet to be determined.

The transcription factors *Tbx4* and *Tbx5* are also present in the limb field, prior to bud formation. However unlike *SnR* or *twist, Tbx4* and *Tbx5* show complimentary expression patterns (Gibson-Brown et al., 1996). Whereas *Tbx5* transcripts are only detected in the forelimb field, *Tbx4* are found almost exclusively in the hindlimb field. These complimentary expression patterns were originally interpreted as an indication that *Tbx4* and *Tbx5* were involved in the specification of hindlimb, and forelimb fates respectively (Ruvinsky and Gibson-Brown, 2000). Subsequent functional analysis has confirmed this original conclusion (Isaac et al., 1998). Leg-to-wing and wing-to-leg tissue mesenchymal tissue grafts, despite their change of position, retain not only their original morphological identity but also their initial *Tbx-4* or *Tbx-5* expression as well. Additionally ectopic expression of either *Tbx4* or *Tbx5* (Takeuchi et al., 1999) demonstrate that *Tbx4* can induce

leg like structures in the wing, and *Tbx-5* can induce wing like structures in the leg. In addition to specifying limb fate it is also thought that the *Tbx* genes are essential for limb outgrowth, as removal of *Tbx-5* activity in zebrafish results in a complete loss of the pectoral fin (Ahn et al., 2002).

1.2.1 FGF Signalling

Although it seems likely that *SnR, twist, Tbx-4* and *Tbx-5* are all necessary for establishing the limb field; they are not sufficient for outgrowth of the limb. As has already been mentioned, for the proximal-distal axis of the limb to develop an AER must be formed (Johnson and Tabin, 1997). Current ideas of AER induction centre on the role of a regulatory loop between two members of the FGF superfamily; FGF-8 and FGF-10 (Martin, 1998). Due to their expression patterns and the fact that beads soaked in FGF and placed into the flanks of embryos are sufficient to induce the formation of ectopic AERs and complete limbs (Cohn et al., 1995).

At the time of limb initiation in the chick embryo, *FGF-8* is expressed in the IM, adjacent to the both the forelimb and hindlimb fields (Vogel et al., 1996). FGF-10 is initially widely expressed in the segmental plate, IM and LPM, but just before the onset of bud formation its expression becomes restricted to the LPM of the limb fields (Ohuchi et al., 1997), this is thought to be due to the FGF-8 expression in the IM (Martin, 1998). Following restriction of its expression FGF-10 is then thought to activate FGF-8 expression in the surface ectoderm of the limb field within a broad strip of cells destined to become the AER (Ohuchi et al., 1997). FGF-8 within the ectoderm is then believed to signal back to the mesoderm to maintain FGF-10 expression throughout development, thus completing the FGF regulatory loop (Martin, 1998).

Recent work examining the role of the *Wnt* genes in AER induction in the chick has suggested that several of the *Wnts* signalling through β-catenin act to mediate the FGF loop (Kawakami et al., 2001). *Wnt-2b* expression was detected in a pattern similar to FGF-8 in the

anterior of the IM (close to the wing field), and ectopic expression of *Wnt-2b* in the flank resulted in the activation of LPM FGF-10, and an ectopic limb. *Wnt-8c* was shown to have a similar limb inducing ability as *Wnt-2b* and to closely mirror FGF-8 expression in the posterior of the IM (close to the leg field). Inhibition of *Wnt-2b* and *Wnt-8c* activity by antagonising the receptor β-catenin, resulted in severe defects in limb outgrowth. A further *Wnt, Wnt-3a* was also implicated in mediating the FGF loop, this time between FGF-10 in the LPM and FGF-8 in the surface ectoderm. Inhibition of β-catenin later in development (just after the initiation of FGF-10 in the LPM) resulted in a down-regulation of FGF-8 in the surface ectoderm and truncated limb structures.

Currently an expanded model of AER induction can be envisaged (see Fig 1.5). Prior to limb bud initiation FGF-10 is expressed in a wide region within the LPM, without any specificity to the presumptive limb fields. At limb bud initiation FGF-10 becomes restricted to the limb fields, by the action *Wnt-2b* and *Wnt-8c* mediated by FGF-8 in the IM. FGF-10 is then believed to stimulate expression of FGF-8 in the presumptive AER, via Wnt-3a. FGF-8

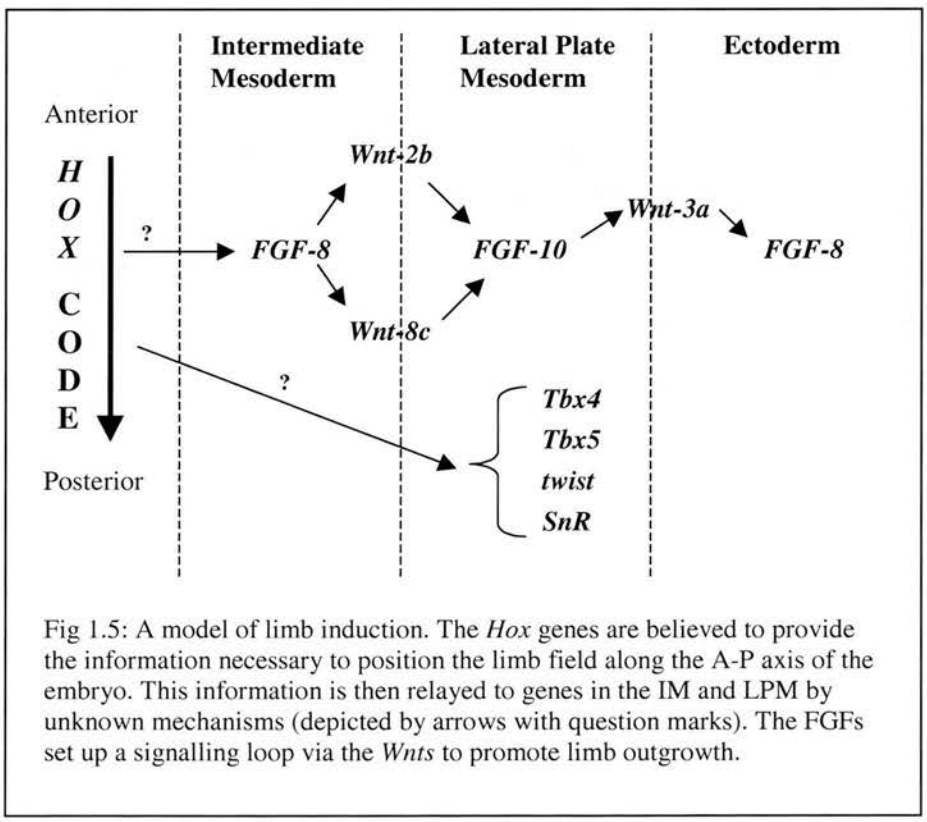


Fig 1.5: A model of limb induction. The *Hox* genes are believed to provide the information necessary to position the limb field along the A-P axis of the embryo. This information is then relayed to genes in the IM and LPM by unknown mechanisms (depicted by arrows with question marks). The FGFs set up a signalling loop via the *Wnts* to promote limb outgrowth.

then completes the loop by signalling back to the LPM to maintain FGF-10 expression.

Although attractive, this model still contains some unresolved issues. Firstly conflicts remain over the role of the IM in the induction of limb development. The model suggests FGF-8 in the IM regulates FGF-10 expression in the LPM via *Wnt* signalling, but this clearly conflicts with work from Fernandez-Teran *et al.* (Fernandez-Teran et al., 1997). They found that by blocking the inductive interaction between the Wolffian duct and the IM rostral to the limb field, they were able to induce apoptosis in the limb field IM. Despite a lack of IM *FGF-8* expression, limb buds still formed suggesting that FGF-8 from the IM is not necessary for initiation of the limb field. Secondly although thought to regulate ectodermal expression of FGF-8 expression in the chick, *Wnt-3a* transcripts cannot be detected in the mouse limb ectoderm (Kawakami et al., 2001). Though $Lef1^{-/-}$ $Tcf^{-/-}$ (transcription factors implicated in the transduction of *Wnt* signals) mice do exhibit a limb phenotype consistent with the downregulation of *Wnt-3a* in chick (Galceran et al., 1999), suggesting that a *Wnt* other than *Wnt-3a* may be acting to upregulate FGF-8 in the presumptive AER. Finally it is still unclear how the FGF/*Wnt* signalling loop interacts with the transcription factors *SnR, twist, Tbx-4* and *Tbx-5*. Even though FGF-10 transcripts become localised to the presumptive limb regions just prior to *SnR* and *twist* expression, beads laced with FGF-10 do not rapidly induce either *SnR* or *twist* when placed in the LPM (Isaac et al., 2000). Additionally FGF-10 knockout mice, still exhibit *Tbx-4* and *Tbx-5* expression (Sekine et al., 1999). Thus it seems unlikely that the FGF signalling loop directly regulates *SnR, twist, Tbx-4* or *Tbx-5*.

1.3 The Dorsal-Ventral Axis and the AER

The section above established that signalling from the underlying mesoderm mediated by FGFs and *Wnts* is essential for limb initiation. However a further question remains, how is the AER formed at a defined location with respect to the D-V axis of the embryo? Fate-mapping studies (Altabef et al., 1997) have revealed that while the mature

AER is a narrow band of cells, located between the dorsal and ventral limb bud ectoderm; it is derived from a relatively broad area of the limb field. Due to the fact that the AER ends up being localised at the interface between dorsal and ventral cells of the limb bud ectoderm. The suggestion has been made that the specification of the D-V axis of the limb and the localisation of the AER are linked by a common mechanism (Meinhardt, 1983).

It is believed that the limb field initially establishes D-V polarity before the emergence of the limb buds, as the grafting of limb field mesenchyme in an opposite D-V orientation leads to the outgrowth of limbs with reversed D-V polarity (Saunders and Ruess C.R., 1974). Juxtaposition of chick wing fields, with two rows of somites, induces bi-dorsal limbs with normal AERs (Michaud et al., 1997), suggesting that not only does a dorsalising signal originate from the somites, but also that the early determination of D-V polarity in the mesoderm of the limb field is not essential for AER formation. However, as limb development progresses, signals that determine D-V polarity are believed to become located within the limb bud ectoderm, as recombination of limb bud mesenchyme within a D-V reversed limb bud ectoderm during bud outgrowth results in the reversal of limb D-V polarity (Geduspan and MacCabe, 1987). Suggesting that prior to limb bud outgrowth the origin of signals determining the D-V polarity of the limb switches from the somites to the ectoderm.

1.3.1 Molecular Cues that Specifiy Limb D-V Polarity

Several genes are expressed along the D-V border of the limb ectoderm, and have been linked to the establishment of D-V fate (Altabef and Tickle, 2002). *Wnt-7a* (Parr and McMahon, 1995) and *Lmx-1* (Vogel et al., 1995) are both restricted to the dorsal limb field prior to and during limb bud outgrowth, *Wnt-7a* is expressed in the ectoderm, *Lmx-1* in the lateral plate mesoderm. Ectopic expression of *Wnt-7a* or *Lmx-1* results in the dorsalising of ventral mesoderm (Riddle et al., 1995), and loss of *Wnt-7a* results in the ventralisation of limbs (Parr and McMahon, 1995). Another gene *Engrailed-1* (*En-1*) is expressed solely in

the ventral ectoderm, and loss of *En-1* results in the dorsalisation of the limbs (Loomis et al., 1996). Together these observations have been linked in the following model (Johnson and Tabin, 1997). *Wnt-7a* is presumed to act as a dorsalising signal from the limb ectoderm, resulting in expression of *Lmx-1* in the underlying mesoderm. *En-1* acts to repress *Wnt-7a* from the ventral ectoderm, thus ensuring that *Lmx-1* is restricted to the dorsal limb mesenchyme. However, although it is reasonable to suppose that the somitic mesoderm is involved in setting up the ectodermal expression patterns of *En-1* and *Wnt-7a*, the molecular signals involved in this transfer of D-V polarity from the mesoderm to the ectoderm remain unknown.

1.3.2 Specifying the AER

Specification of the AER is believed to involve a member of the *fringe* family, *Radical fringe* (*r-Fng*). *r-Fng* has been shown to be expressed in the dorsal ectoderm of chick limb buds before and during limb bud outgrowth with the highest expression observed in the AER (Laufer et al., 1997; Rodriguez-Esteban et al., 1997). Ectopic but localised expression of *r-Fng* in the ventral limb bud results in the induction of small AER regions in the ventral limb bud, which go on to develop biventral extra digits. Uniform expression of *r-Fng* in the ventral limb bud partially or completely suppresses endogenous AER formation (Laufer et al., 1997; Rodriguez-Esteban et al., 1997). These dramatic results have been interpreted to mean that the AER forms at the interval between *r-Fng* expressing and non-expressing cells (Zeller and Duboule, 1997). A remarkable parallel with invertebrate development as *Drosophila fringe* is believed to play a similar role in formation of the wing margin(Kim et al., 1995).

To specify the AER correctly in the developing limb bud *r-Fng* has to be restricted to the dorsal ectoderm of the limb bud, this appears to be achieved through the activity of *En-1* (Laufer et al., 1997) which as previously mentioned is expressed in the ventral ectoderm. Ectopic expression of *En-1* in dorsal chick limb buds represses transcription of *r-*

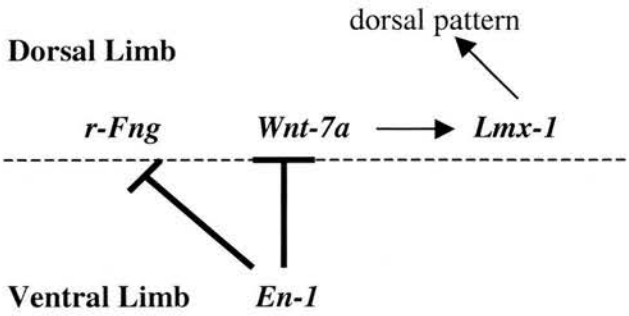


Fig 1.6: A diagram depicting the gene interactions that act to position the AER and pattern the limb with respect to the D-V axis. *En-1* restricts *r-Fng* and *Wnt-7a* expression to the dorsal ectoderm. The AER forms at the boundry of *r-Fng* expressing and non-expressing cells. *Wnt-7a* triggers expression of *Lmx-1* in the mesoderm, which promotes a dorsal fate in the limb bud.

Fng resulting in the formation of ectopic AERs and the formation of bi-dorsal ectopic digits (Laufer et al., 1997) (Logan et al., 1997). Thus specification of the AER in the chick appears to rely on an antagonistic relationship between *r-Fng* and *En-1*, which positions the AER at the D-V boundary of the developing limb ectoderm due to their respective expression patterns.

The ectopic expression of *r-Fng* and *En-1* suggest that the formation of an AER does not necessarily have to occur at the boundary between dorsal and ventral cells. As AERs can be induced in solely dorsal or ventral tissue, resulting in digits without the normal D-V patterning. This is further supported by the chick mutation Eudiplopodia (Goetinck P.F., 1964; Carrington and Fallon, 1986), which displays ectopic AERs that do not form at the interface of dorsal (*Wnt-7a* expressing) and ventral (*En-1*) expressing cells. Instead they form within the dorsal ectoderm resulting in dorsalised limbs. Additionally *Wnt-7a* homozygous mutant mice still form correctly placed and functional AERs, but develop ventralised limbs. So although at first glance it would seem that the mechanisms involved in patterning the D-V axis of the limb and the positioning of the AER are tightly linked, it is now obvious that they are in fact separable. A summary diagram depicting the interaction of *En-1* with *r-Fng* and *Wnt-7a* is shown in Fig 1.6.

Although the juxtaposition of *r-Fng* expressing and non-expressing cells provides an excellent model for the localisation of the AER in the chick. The mouse *r-Fng* gene is not required for limb development, as its removal by homologous recombination does not exhibit a limb phenotype (Moran et al., 1999). The reasons for this are unclear though it has

been suggested that a functional overlap with other members of the *fringe* family could exist in mice (Moran et al., 1999). Combined knockouts of multiple *fringe* members may resolve this issue.

1.4 The Proximal-Distal Axis of the limb

1.4.1 Outgrowth of the limb

As noted above the AER is essential for growth along the P-D axis of the limb, as it is believed to provide molecular signals to the underlying mesenchyme that promote limb outgrowth (Rowe and Fallon, 1982). Experiments have shown that upon removal of their AER, limbs can be rescued by the application of beads soaked in FGF to their distal tips, leading to the suggestion that FGFs are the AER derived signal that leads to outgrowth of the limb (Fallon et al., 1994; Niswander et al., 1993). Four of the 22 known FGF genes, *Fgf4*, *Fgf8*, *Fgf9* and *Fgf17*, display AER specific expression domains (Martin, 1998), and a further two *Fgf2* and *Fgf10* are expressed in the mesenchyme underlying it (Martin, 1998). Although it would seem that any of the above FGFs can rescue a limb devoid of an AER, strangely, not all the above FGFs are essential for limb outgrowth. The individual loss of *Fgf2* , *Fgf9* (Martin, 1998), or *Fgf17* (Xu et al., 2000)has no effect on limb development.

Genetic analysis of *Fgf4* and *Fgf8* function in the limb has required a tissue-specific inactivation approach, as null homozygosity in either gene cause embryonic lethality before limbs develop (Feldman et al., 1995; Meyers et al., 1998). Individual inactivation of *Fgf4* does not elicit a limb phenotype (Moon et al., 2000), but removal of *Fgf8* activity in the limb results in hypoplasia of all three limb segments (Lewandoski et al., 2000; Moon and Capecchi, 2000). Recent double conditional knockouts of *Fgf4* and *Fgf8* result in the failure of limbs to develop(Sun et al., 2002). Thus although there is considerable functional redundancy between the FGFs *Fgf4* and *Fgf8* appear crucial for limb outgrowth. The major function of the FGFs in the AER is believed to be the stimulation of cell proliferation within the developing limb, to ensure that there are enough cells to form all the required limb

structures. Studies employing markers for cell proliferation have directly demonstrated that FGFs are mitogens for limb bud mesenchyme (Dealy et al., 1997), and when *Fgf4* and *Fgf8* are switched off part way through bud outgrowth, the resulting limb is much smaller than normal (Sun et al., 2002).

As mentioned previously *Fgf10* is thought to play a vital role in promoting the initial outgrowth of the AER by establishing a positive feedback loop with *Fgf8* (Martin, 1998). Continued expression of *Fgf10* in the limb bud mesenchyme is thought to be due to this same feedback loop and be required to maintain FGF expression in the AER. *FGF10* has been suggested as the "AER maintenance factor" produced by the distal limb mesenchyme, the existence of which has been inferred from classical embryology experiments (Martin, 1998). Although essential for generating the cells necessary for the creation of the P-D axis, other mechanisms are believed to be required to pattern the P-D skeletal elements.

1.4.2 Patterning the Proximal-Distal axis of the limb

Traditionally theories concerning the patterning of the P-D axis have been dominated by the progress zone model (PZ model), first suggested over 30 years ago in response to AER ablation experiments (Summerbell D.Lewis J.H.Wolpert L., 1974). It was observed that following the removal of the AER the limb becomes truncated at a proximal-distal level, that was dependent on the stage of the limb bud at the time of AER removal. In other words removal of the AER early in development resulted in a limb containing only proximal structures, whereas late removal of the AER resulted in loss of only the most distal limb elements. Additionally it was noted that exchange of AERs between young and old limb buds did not disrupt the normal proximal-distal patterning of skeletal elements. Thus the information required to pattern the P-D axis was contained within the bud mesenchyme not the AER(Summerbell D.Lewis J.H.Wolpert L., 1974).

According to the progress zone model these observations could be explained if, cell fate along the P-D axis is determined by the amount of time spent in the progress zone, a region of distal undifferentiated mesenchyme located just underneath the AER (see Fig 1.7).

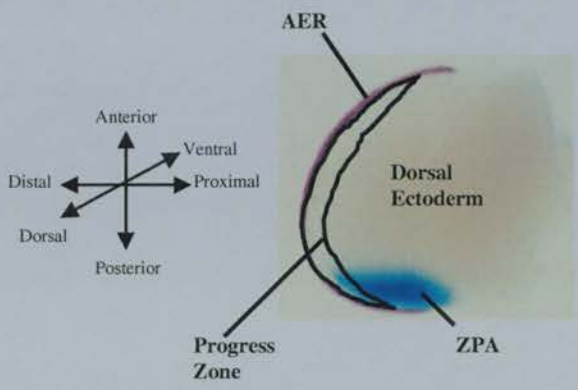


Fig 1.7 E 10.5 mouse limb bud, the location of the progress zone is highlighted.

Cells exiting the zone after a short time acquire a proximal positional address, ultimately forming proximal limb structures such as the humerus. Whereas cells remaining in the progress zone for longer gain ever more distal fates, such that the last cells to leave gives rise to the most terminal structures such as the phalanges. The model predicts that distal mesenchyme cells measure the length of time they are within the progress zone, possibly by exhibiting a quantitative response to a factor that accumulates in response to ridge signals. Thus removal of the AER stops the progress zone from accumulating more factor, prematurely ending P/D patterning. Replacement of the AER even from a limb at a different stage enables the progress zone to begin accumulating more AER response factors again, and P-D patterning continues from the same place. Several transcription factors such as *Lhx2, Msx1, Evx1,* and *Slug* have been suggested as possible candidates for the accumulating factor, (Rodriguez-Esteban et al., 1998; Niswander and Martin, 1993; Davidson et al., 1991; Ros et al., 1997) as all have appropriate expression patterns and are regulated by FGF signalling in the AER.

1.4.3 The Progress Zone Model Debunked?

Recent data has thrown doubt on the validity of the progress zone model. Dudley *et al.* (Dudley et al., 2002) examined the patterns of cell death and proliferation in limb buds after the removal of the AER. In confirmation of earlier reports (Rowe et al., 1982), they

found that 6 to 8 hours after ablation of the AER, the apical mesenchyme undergoes reduced proliferation and substantial cell death stretching through a zone approximately 200 micrometers proximally from the AER. The extent of the tissue lost after AER removal corresponds to the amount of distal deletion seen, however swift replacement of the AER after removal prevents cell death and maintains apical proliferation.

In addition Dudley *et al.* used the lipophilic dyes DiI and DiO to examine the fates of cells marked early in limb development. They found that in early stage chick wing buds (stage 19) cells directly beneath the AER appeared only in the autopod, those marked 100 micrometers proximal to the AER were found only in the zeugopod and those labelled 200 micrometers below the AER were found in the stylopod. The distal part of a stage 19 wing bud, comprising the AER and the presumptive (as labelled) autopodial mesoderm was then grafted to the stump of an amputated leg bud. The graft formed only digit-like skeletal elements, i.e. only autopodial structures, these elements exhibited the wing-specific *Tbx5* marker and the autopodial specific marker *Hoxa-13*. Therefore, the grafted tissue was specified as autopod at stage 19, a much earlier stage than is predicted by the PZ model. The loss of distal structures upon removal of the AER in stage 19 chick limbs could now be explained by the loss of autopod and zeugopod cell progenitors in the region of cell death stretching 200 micrometers proximally from the AER.

Further data inconsistent with the PZ model has recently been revealed in double *Fgf4* and *Fgf8* conditional knockout alleles created by *Sun et al.*(Sun et al., 2002). These mice fail to develop hindlimbs, but forelimbs develop skeletal elements from all three proximal-distal segments due to transient expression of *Fgf4* and *Fgf8* in their AERs. However after inactivation of *Fgf4* and *Fgf8* the forelimbs were substantially diminished in size by cell death in the proximal region. The authors suggest that this cell death reduces the number of skeletal progenitors available to form the forelimb. In these mice, proximal regions (zeugopodial) elements were invariablly deficient when distal (autopodial) elements were present. The PZ model predicts that specification of proximal elements should be

complete before the specification of more distal elements is initiated. This is clearly not the case in the *Fgf4* and *Fgf8* conditional mice.

The results of Dudley *et al.* and *Sun et al.*, indicate that the proximal-distal axis of the limb is specified much earlier than originally predicted by the PZ model. The presence of distinct autopodial progenitors in stage 19 chick wings, suggests that the P-D axis is specified as a pre-pattern, i.e. cells are designated a P-D fate early in limb bud outgrowth, or perhaps even in the lateral plate mesoderm before the limb buds appear. The establishment of such a pre-pattern would obviously depend on differential gene expression between the progenitors of the stylopod, zeugopod and autopod. Excellent candidates for possible mediators of a P-D pre-pattern are the homeobox genes *Meis-1*, *Meis-2* and *Pbx-1*. *Meis-1* and *Meis-2* are believed to activate expression of *Pbx-1* (Mercader et al., 1999) and are expressed throughout the lateral mesoderm, and the areas of the limb bud that become the stylopod. Over-expression of *Meis-1* in the chick limb results in a reduction in size and proximilisation of the zeugopod and autopod (Mercader et al., 1999). *Meis-1* and *Meis-2* are activated by retinoic acid (RA) present in the limb bud and repressed by *FGF* signalling from the AER (Mercader et al., 2000). Thus, it is feasible that *Meis-1* and *Meis-2* act to pre-pattern the stylopod by being restricted to the proximal limb bud by *FGF* signalling from the AER.

Whatever the precise nature of the P-D pre-pattern it is probable that it does not fix cell fates irreversibly. In addition to their autopod progenitor grafts Dudley *et al.* also dissociated apical limb mesenchyme cells from wing buds at stage 20, reassembled them in ectodermal hulls, and allowed them to develop. The recombinants formed skeletal elements from all three limb sections. However, recombinants formed from apical cells of later stage wings formed progressively less proximal structures. Therefore it would appear that initially cells are divided into P-D progenitor populations by a P-D pre-pattern, but only establish a fixed fate later in development.

1.5 The Anterior Posterior Axis and the ZPA

As has already been mentioned, surgical manipulation of the chick embryo resulted in the identification of the ZPA, the area of the limb bud believed to be responsible for patterning the A-P axis of the developing limb bud (Saunders, 1968). Soon after the identification of the ZPA Lewis Wolpert proposed a mechanistic model for its function (Wolpert L.Macpherson I.Todd I., 1969). Wolpert realised that to generate a complex structure such as a limb bud, embryonic cells must know where they are in relation to the developing structure, i.e. they must have positional information. Using the analogy of generating the pattern of a French flag (one third blue, one third red and one third white) Wolpert devised a positional information paradigm that relied on a concentration gradient. Wolpert envisaged that a chemical (a morphogen) produced on one side of a field (such as in the ZPA) would set up a concentration gradient across the field. Near to its point of origin the morphogen would be at a high concentration, but the other end of the field would receive a low concentration of the morphogen. Cells would then differentiate according to the concentration of the morphogen they receive: for example blue at high concentrations, white at intermediate, and red at low concentrations (see Fig 1.8).

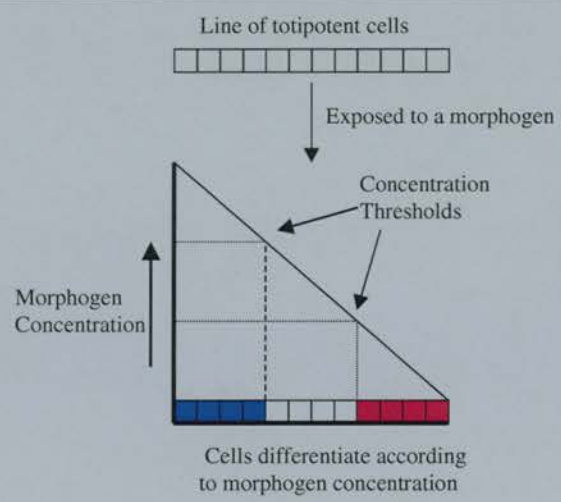


Fig 1.8: Wolpert's French Flag model. A line of totipotent cells detects the concentration of a particular morphogen within a developmental field and differentiates accordingly. High concentrations result in blue cells, intermediate concentrations in white cells and low conetrations in red cells. Thus the pattern of a French flag is formed.

Several lines of evidence support a gradient model for ZPA function. Firstly the number and morphology of ectopic digits formed when cells from the ZPA are transplanted into the anterior of the limb is proportional to the number of ZPA cells transplanted (Tickle, 1981). A large number of cells

results in a full mirror image duplication of the digits, 4-3-2-2-3-4 (chick wings have three digits, which are designated from the posterior to the anterior as 4-3-2). However transplantation of a smaller numbers of cells result in progressively less ectopic digits 3-2-2-3-4, or 2-2-3-4. Secondly, attenuation of the endogenous ZPA's activity, by reducing the number of cells within the posterior limb bud mesenchyme, reduces the number of endogenous digits proportionally (Smith et al., 1978). Both results suggest that the ZPA produces a factor that acts in a concentration dependent manner. Finally, transplantation of ZPA cells to the apex, as opposed to the anterior, of a limb bud results in the formation of a complex set of digits, consistent with the superposition of an ectopic and endogenous concentration gradient (Tickle et al., 1975).

1.5.1 Sonic hedgehog the morphogen?

The molecule identified as the morphogen produced by the ZPA is Sonic hedgehog (Shh), an orthologue of the Drosophilia gene hedgehog (hh) (Riddle et al., 1993). Shh has been shown to co-localise with the ZPA in mouse and chick limb buds (Riddle et al., 1993), and its expression can be induced by retinoic acid (RA) a known inducer of polarising activity (Riddle et al., 1993). Further proof came from the grafting of Shh producing cells or beads impregnated with SHH protein (Chang et al., 1994; Lopez-Martinez et al., 1995) into the anterior of limb buds, both were capable of producing mirror image digit duplications. The removal of Shh in knockout mice also reveals that Shh has a polarising function (Chiang et al., 1996). Mice homozygous for the Shh knockout allele (hereafter referred to as Shhnull / null mice) exhibit severe defects in many embryonic structures including the limb (Chiang et al., 2001). Mutants have four limbs with recognisable A-P patterning in the stylopod (humerus/femur), but below the stylopod the A-P patterning is severely disrupted. Skeletal elements of the zeugopod form but exhibit defects in A-P patterning. The forelimb autopod, is represented by a single distal cartilage element whereas the hindlimb autopod consists of a single digit 1 (mice hindlimbs have five digits designated 5-4-3-2-1 from the posterior to the

anterior). Thus it would seem that Shh is vital for the activity of the ZPA, and for A-P patterning the zeugopod and autopod of the limb.

However, to act as a morphogen as proposed by Wolpert's gradient model SHH must be shown to diffuse from the ZPA and be able to set up a concentration gradient across the limb bud. The currently available evidence of whether or not SHH can act as a morphogen is conflicting. Although the grafting of Shh producing cells/beads induces digit-duplications in a dose dependent manner, a large number of studies have failed to detect SHH from its site of synthesis in the ZPA (Johnson and Tabin, 1997). Additionally, when expressed in insect and mammalian cell lines both HH and SHH are tightly bound to the cell surface (Chang et al., 1994; Lopez-Martinez et al., 1995; Lee et al., 1994; Bumcrot et al., 1995). The reason for this adherence was believed to be due to the post-translational modification of the SHH pro-peptide. The pro-peptide is cleaved to release an active 19kDa N-terminal fragment (N-Shhp) which is retained at the cell surface by the attachment of cholesterol to its C terminus (Porter et al., 1996).

However, more recent data has suggested that the addition of cholesterol to N-Shhp may in fact be essential for long range signalling of *Shh*. Lewis *et al.*(Lewis et al., 2001) used gene targeting in mouse embryonic stem cells to terminate translation immediately downstream of the glycine residue at position 198 of the Shh peptide. This recombinant allele (*N-Shh*) generated a truncated Shh protein differing from the autoproteolytically cleaved wild-type protein by the absence of the cholesterol. To determine the biological significance of the cholesterol modification of SHH peptide, the *N-Shh* allele was crossed onto a Shh^{null} background to generate *N-Shh*/Shh^{null} mice. Thus the only SHH protein generated by these mice did not exhibit the cholesterol modification.

N-Shh/Shh^{null} mice demonstrated normal outgrowth of the limbs and three digits formed (5-4-1), as opposed to the severely truncated limbs exhibited by $Shh^{null\,/\,null}$ mice. This result has been interpreted to mean that the addition of cholesterol may actually increase the range of *Shh* signalling (Lewis et al., 2001), as the *N-Shh*/Shh^{null} mice display a lack of SHH

mediated A-P patterning whereas $Shh^{null/+}$ are known to have a wild type phenotype (Chiang et al., 1996). Another group has suggested a mechanism by which cholesterol may facilitate the diffusion of SHH across the limb bud. Zeng *et al.* (Zeng et al., 2001) identified a naturally occurring, freely diffusible form of Shh (s-ShhNp) that is cholesterol modified and multimeric. Their work confirmed that s-ShhNp was biologically potent, and could be isolated from chick limb bud tissue, large amounts from the posterior region with decreasing amounts towards the anterior of the bud. Though it should be noted that Zeng *et al.* were not able to rule out the possibility that additional factors may have produced the low levels of SHH in the assay used to detect s-ShhNp in the anterior of limb buds. Zeng *et al.* suggested a model whereby ShhNp concentrates within the cell membrane, where it multimerizes with is lipid attachments sequestered to the interior of the multimer to form soluble s-ShhNp. s-ShhNp then diffuses from its site of synthesis to provide long-range patterning information (i.e. act as a morphogen). The ability of s-ShhNp to work at very low concentrations may explain why previous attempts to detect a gradient of SHH across the limb bud using immunohistochemistry have failed.

The most recent evidence suggests that SHH can diffuse a substantial distance from the ZPA, however there are several mechanisms that are known to restrict the movement of SHH. In *Drosophila* and presumably also in vertebrates, HH diffusion is contained by binding to its receptor *Patched* (*ptc*) (Chen and Struhl, 1996), and another more recently discovered cell surface protein, Hip (Hedgehog interacting protein) (Chuang and McMahon, 1999). *Hedgehog* signalling induces high-level expression of the genes encoding these proteins, thus the signalling molecule itself attenuates and limits its own action. The reason for this remains unknown, though in the polydactylous chick mutant *talpid* there is no high level expression of *ptc* and this has been suggested to result in a wider diffusion of Shh protein, ultimately causing the formation of extra digits (Caruccio et al., 1999). Thus the negative regulation of Shh diffusion may also be important for the correct regulation of A-P axis.

1.5.2 Interpreting the A-P concentration gradient of Shh

The most recent data does seem to support a role for *Shh* as a morphogen, but the method by which a concentration gradient of SHH could be interpreted is only now being elucidated. One known target of *Shh* signalling is the transcription factor *Gli3* the orthologue of the *Drosophila* gene *cubitus interuptus* (*Ci*) (Methot and Basler, 2001). In the absence of *hh* signalling *Ci* (Ci155) is processed to generate a 75 kDa amino-terminal fragment (Ci75) that acts to repress transcriptional targets genes (Aza-Blanc et al., 1997). The presence of HH prevents the generation of Ci75, and also causes Ci155 to act as a transcriptional activator (Methot and Basler, 1999). *Shh* is believed to act in a similar way on *Gli3*, inhibiting the formation of Gli3 repressor (Gli3R), while promoting the formation of Gli3 activator (Gli3A) (Ingham and McMahon, 2001). Interestingly in chick and mouse limb buds A-P concentration gradients of Gli3R and Gli3A are observed, the posterior of the limb exhibits high levels of Gli3A but low levels of Gli3R, the converse is true in the anterior of the limb (Wang et al., 2000).

Recent analysis of *Gli3/Shh* compound mutants has revealed an intriguing limb phenotype (Litingtung et al., 2002). As has already been mentioned $Shh^{null/null}$ mice exhibit severe limb deformities, including A-P skeletal deficiencies (Chiang et al., 1996). The semidominant mouse mutation *Extra toes* (Xt^J) generates a *Gli3* null allele, hereafter referred to as $Gli3^{null}$. $Gli3^{null}/Gli3^{null}$ limbs are normal down to the level of the wrist/ankle, but exhibit severe polydactyly 6 to 11 digits (Hui and Joyner, 1993), which traditionally has been attributed to ectopic *Shh* expression in the anterior of the limb bud (Masuya et al., 1995; Masuya et al., 1997). However $Gli3^{null}/Gli3^{null}$ limbs exhibit complete loss of digit identity, a result that has always puzzled researchers as anterior ectopic expression of *Shh* usually results in mirror-image duplications of digits with wild type identities (Masuya et al., 1995; Masuya et al., 1997).

The phenotype of $Shh^{null/null}$ $Gli3^{null/null}$ limbs is virtually identical to that of $Gli3^{null/null}$ limbs (Litingtung et al., 2002), suggesting that *Shh* must act through *Gli3* for normal ZPA function. $Shh^{null/null}$ $Gli3^{null/+}$ limbs exhibited three or four digits, an intermediate phenotype to the single digit of $Shh^{null/null}$ mice or the 6-11 digits of $Shh^{null/null}$ $Gli3^{null/null}$. Additonally all the digits generated by $Shh^{null/null}$ $Gli3^{null/+}$ mice are identifiable as digit 1 regardless of their position along the A-P axis. Therefore, although reducing *Gli3* dosage in $Shh^{null/null}$ can mimic *Shh* signalling in rescuing digit formation, posterior digit identities are not restored; the regulation of digit number and identity are *Gli3* dependant but seperable processes. The most likely scenario is that Shh acts to pattern the A/P of the autopod by setting up a gradient of Gli3A:Gli3R. Gli3A can be envisaged as promoting a posterior identity in digits, whereas Gli3R can be viewed as inhibiting digit formation (See Fig 1.9). In $Shh^{null/null}$ limbs only Gli3R is present (due to a lack of SHH) therefore digit formation is inhibited and the single digit that remains takes on an anterior identity. However in $Gli3^{null/null}$ neither Gli3A or Gli3R is present, therefore no A-P digit identity is established and a polydactylous phenotype is observed as Gli3R is not available to restrict the number of digits to five.

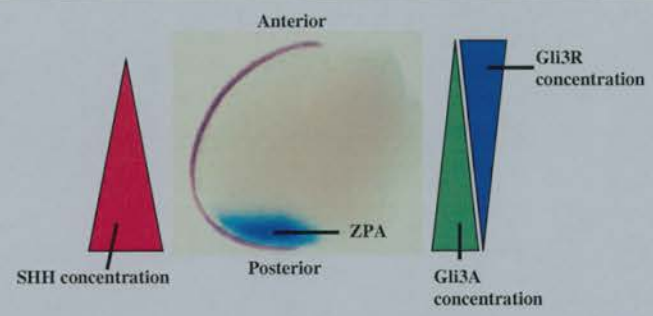


Fig 1.9: A concentration gradient of SHH protein across the A-P axis of the limb is predicted to set up reciprocal conentration gradients of Gli3A and Gli3R. Such that towards the posterior of the limb there are high levels of Gli3A, but towards the anterior there are high levels of Gli3R.

1.6 Pre-patterning the limb field

Great strides have been made in understanding the patterning of the A-P axis of the autopod. However the same cannot be said of the zeugopod and stylopod. Mice null for *Shh* and *Gli3* exhibit no zeugopod or stylopod abnormalities (Litingtung et al., 2002), suggesting that mechanisms other than a concentration gradient of Gli3A:Gli3R across the A-P axis of

the limb specifies their A-P polarity. Currently these mechanisms are unknown, though determination of zeugopod and stylopod A-P polarity could be part of a limb field pre-patterning event. Such an event can be defined as, "the establishment of limb field positional information with regard to the limb's eventual three axis prior to limb bud formation". Evidence for the formation of a limb pre-pattern has already been suggested for the D-V and P-D axis of the limb. *Wnt-7a* (Parr and McMahon, 1995)and *En-1* (Logan et al., 1997) are expressed in dorsal and ventral ectodermal domains respectively prior to limb bud outgrowth, and the *Meis* genes have been suggested to play a role in the establishment of stylopod identity prior to bud initiation (Mercader et al., 1999).

At present there are few candidate genes for establishing an A-P pre-pattern. *Gli3* and *Alx4* are both expressed in anteriorly restricted domains in the pre-limb bud mesenchyme, but the removal of either gene's activity only results in A-P defects in the autopod (Hui and Joyner, 1993; Qu et al., 1998). Thus although possibly important in establishing an anterior pre-pattern for the autopod it seems unlikely that *Gli3* and *Alx4* are involved in establishing an anterior pre-pattern for the zeugopod or stylopod. However, the transcription factor *dHAND* is also expressed before and during the outgrowth of the limb bud, in the posterior domain of the limb field (Charite et al., 2000). Additionally ectopic expression of *dHAND* in the anterior limb field results in the complete mirror-image duplications of posterior elements of both the zeugopod and autopod (Charite et al., 2000). Therefore it is feasible that *dHAND* may play a role in establishing a posterior domain in the limb pre-pattern for both the zeugopod and autopod.

A striking fact is that most limb mutations seen in mouse and chick exhibit patterning defects in the more distal elements of the limb, proximal elements always seem to be less severely affected or normal. Is this because the perturbation of factors essential for the patterning of the stylopod and to a lesser extent the zeugopod affect a limb pre-pattern to such a severe extent that the limb never develops? Although an intriguing possibility it

awaits the identification of further genes exhibiting early limb field expression patterns, and their conditional removal at specific stages of limb development.

1.7 What of the Bone Morphogenetic Proteins?

Another set of proteins that have been suggested to play a role in determining the A-P axis of the developing limb bud are the bone morphogenetic proteins (BMPs) (Cohn and Bright, 1999). *Bmp2* is expressed in a pattern that broadly overlaps that of *Shh* in the limb bud (Francis et al., 1994b), and has been shown to have mild polarising activity (Duprez et al., 1996). *Bmp2* has been suggested to mediate the A-P patterning effects of *Shh*, as *Shh* was not previously believed to be able to diffuse out of the ZPA (Cohn and Bright, 1999), where as the *Bmps* are known to be able to diffuse widely (Cohn and Bright, 1999). However, with the discovery of a freely diffusible form of SHH, s-ShhNp (Zeng et al., 2001), the role of *Bmp2* in the early limb bud is unclear.

However, recent data suggests that the *Bmps* may be involved in specifying A-P digital fates at a much later stage, after *Shh* has disappeared from the ZPA. Work by Dahn and Fallon (Dahn and Fallon, 2000) suggests digital identities are specified by the interdigital mesoderm (IDM), before its regression. More posterior IDM specifies more posterior digital identities, and each digit will develop in accordance to the most posterior cue received. Additionally inhibition of IDM BMPs can transform digital identity, suggesting the BMPs may mediate this process.

As well as playing a role in the A-P patterning of the limb, BMP2, BMP4 and BMP7 have also been implicated in regulating the apoptosis of interdigital mesenchyme. Misexpression of Noggin, an antagonist of BMP signalling, in the ectoderm of transgenic mice results in extensive tissue syndactyly, due to a lack of apoptosis in the interdigital mesoderm (Guha et al., 2002). Additionally, co-expression of BMP4 with Noggin, in double transgenic embryos restores interdigital apoptosis (Guha et al., 2002). Thus, BMP signalling is causally implicated in regulating the regression of interdigital mesenchyme in mammals,

consistent with prior findings in avians (Yokouchi et al., 1996; Zou and Niswander, 1996). However, no significant change in the expression pattern of *Msx1* and *Msx2* was observed in Noggin transgenic limbs (Guha et al., 2002), although *Msx2* has been implicated in BMP mediated interdigital apoptosis in the chick (Zou and Niswander, 1996). The reasons for this are currently unclear.

A role for the BMPs in the patterning of the D-V axis of the limb has also been suggested by the fact that BMP2, BMP4 and BMP7 are expressed in the early chick ventral ectoderm, co-incident with *En-1* (Francis et al., 1994a) (Francis-West et al., 1995). Further evidence comes from the misexpression of Noggin in the chick limb, which results in a partial or total absence of EN1 in the ventral ectoderm, and a dorsalised limb (Pizette et al., 2001). Conversely, misexpression of constitutively activated *BmpRIB* (a BMP receptor) results in misexpression of *En-1* in the dorsal ectoderm, ventralising the limb (Pizette et al., 2001). Combined with the fact that misexpression of *En-1* does not alter BMP2, BMP4 and BMP7 expression (Pizette et al., 2001), this data suggests that the BMPs act as ventralising signals upstream of *En-1*.

1.8 The AER and the ZPA regulate each other

For simplicities sake the patterning of the limb has been portrayed as the definement of three separate signalling regions that then act individually to pattern the three limb axis. However the reality of limb development is obviously far more complex, as the various signalling regions are believed to act in concert with each other to pattern the limb as a whole (Johnson and Tabin, 1997). A good example of this is the interaction between the AER and ZPA. ZPA deletion experiments have revealed that not only is the A-P axis affected but the truncations occur along the P-D axis as well (Johnson and Tabin, 1997). This is further illustrated in $Shh^{null/null}$ mice which exhibit P-D deletions (Chiang et al., 1996; Chiang et al., 2001). The reason for this is that *Shh* is required to maintain the AER, lack of *Shh* expression eventually results in the loss of the AER. In turn FGF signalling from the

AER is believed to maintain the ZPA, as loss of *Fgf8* quickly results in the down-regulation of *Shh* ZPA expression (Martin, 1998).

1.9 Positioning Shh to the ZPA

The discovery of *Shh* as the morphogen released from the ZPA, allows the following question to be asked: what mechanisms lead to restricted *Shh* expression in the posterior limb bud? The ability of tissues to express *Shh* in the forelimb ZPA has been linked to *Hoxb8* expression in the lateral plate mesoderm (Lu et al., 1997). Its endogenous expression correlates well with the future forelimb ZPA and ectopic expression of *Hoxb8* results in the formation of ectopic ZPA tissue (Charite et al., 1994). The realisation that *Hoxb8* may have a role in determining the position of the ZPA links the determination of the A-P polarity of the main body axis with that of the autopod.

However grafting experiments in both the chick and the mouse have revealed that in fact most of the flank to greater or lesser degrees has the capacity to induce a polarising region if placed in a permissive environment (Tanaka et al., 2000). Thus, in addition to a positive determinant of ZPA position, cells along the flank must also be inhibited from expressing *Shh*. Indeed, this is further highlighted by the examination of the *hemimelia-luxate* group of mouse mutants that exhibit pre-axial polydactyly (Masuya et al., 1995; Masuya et al., 1997), (additional digits on the anterior side of the limb) Extra toes (*Xt*), Recombination induced mutant 4 (*Rim4*), Strong's luxoid (*ls*), luxate (*lx*), X-linked polydactyly (*Xpl*) Hemimelic extra toes (*Hx*) dominant hemimelia (*Dh*) (Lettice *et al.* 1999) and sasquatch (*Ssq*) (Sharpe et al. 1999) all exhibit an ectopic anterior domain of *Shh* expression in developing limb buds (see Fig 1.10). These mutants suggest that *Shh* must be actively repressed in the anterior domain of the limb bud. The nature of the repressor mechanism is unknown for the majority of these mutants, with the exception of *Xt* and *ls*, which have been shown to be due mutations that result in non-functional *Gli3* (Hui and Joyner, 1993) and *Alx4* (Qu et al., 1998) respectively. Due to these observations *Gli3* and

Alx4 have been suggested as repressors of *Shh* expression in the anterior of the limb bud. To gain a further insight into the mechanism by which *Shh* expression is directed to the ZPA of developing limbs our group has been examining the *Ssq* mutation.

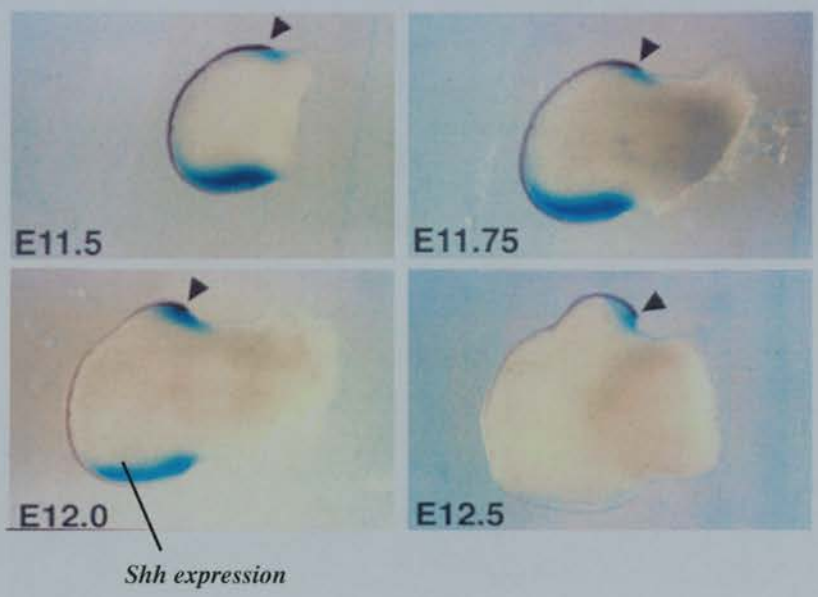


Fig1.10: Limb buds from *Ssq* animals depicting anterior ectopic expression of *Shh* in the limb bud. All the members of the *hemimelia-luxate* group of mutations exhibit similar ectopic *Shh* expression, this ectopic expression is believed to cause pre-axial polydactyly.

1.10 The sasquatch mutation

The sasquatch mutation (*Ssq*) arose as the result of the random insertion of a *Hoxb1 human placental alkaline phosphatase* (*HPAP*) reporter construct, in one of eight transgenic lines (Sharpe et al., 1999). In adults heterozygous for the *Ssq* mutation the forelimbs are normal, but the hindlimbs display preaxial polydactyly (see Fig 1.11). Homozygotes exhibit more extensive limb abnormalities, the hindlimb polydactyly is more severe, and the zeugopod displays hemimelia (reduction in the length of the tibia). Homozygous forelimbs also display preaxial polydactyly and slight hemimelia. The semi-dominant nature of the *Ssq* mutation and the observation that hindlimbs are more severely affected than forelimbs is consistent with other members of the *hemimelia-luxate* group of mutants. However, unlike

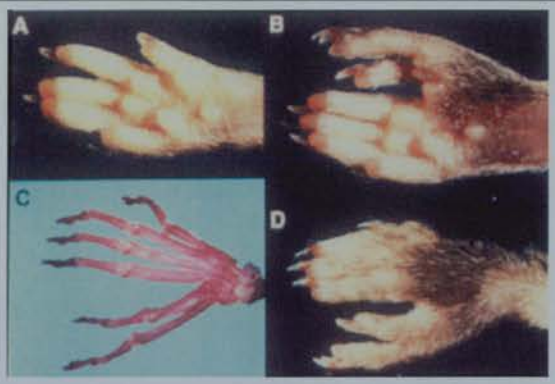


Fig 1.11: Pre-axial polydactyly as displayed by *Ssq* heterozygous mice. **A** is a wild type hindlimb for comparison. **B**, **C** and **D** are all *Ssq* heterozygous hindlimbs (taken from Sharpe *et al.*).

the rest of the *hemimelia-luxate* group, but in common with *Hx*, no other defects are seen outside of the limb in *Ssq* mice.

1.10.1 The Ssq transgene insertion site

In an effort to identify the site of the *Ssq* mutation metaphase nuclei from *Ssq* animals were probed with the HPAP transgene using fluorescence *in situ* hybridisation (FISH), and two insertion sites were located to the proximal half of chromosome 5. *Msx1* and *lx* loci were shown to lie in their normal orientation between the two integration sites indicating that the insertion events had not created a large chromosomal inversion. *Ssq* mice were backcrossed to *Mus musculus castaneus* to establish which integration site colocalised with the mutation. Markers for chromosome 5 exhibited expected recombination frequencies, re-enforcing the idea that the HPAP insertions had not in fact created a large chromosomal rearrangement. The more proximal of the two insertion sites was found to be associate with the mutation, henceforth called the *Ssq* insertion. The physical location of the *Ssq* insertion was established on interphase spreads, which demonstrated that it had integrated approximately 800kb away from *Shh*.

A more precise location for the *Ssq* insertion was established by Dr Laura Lettice (MRC, Human Genetics Unit, Edinburgh) using a λ genomic library made from *Ssq*/*Ssq* mice (Lettice et al., 2002). Clones were isolated using the HPAP transgene as a probe, and a clone incorporating a junction between the transgene and surrounding genomic DNA was identified. The genomic section of this λ clone was then used to identify mouse PAC clones. Subsequent exon trapping of the mouse PAC clones identified a surrounding gene, *Lmbr1*. Further elucidation of the insertion site using a cosmid library revealed that the *Ssq* insertion

consisted of multiple HPAP insertions and had integrated into intron 5 of the *Lmbr1* gene duplicating 20kb of genomic DNA in the process (see Fig 1.12).

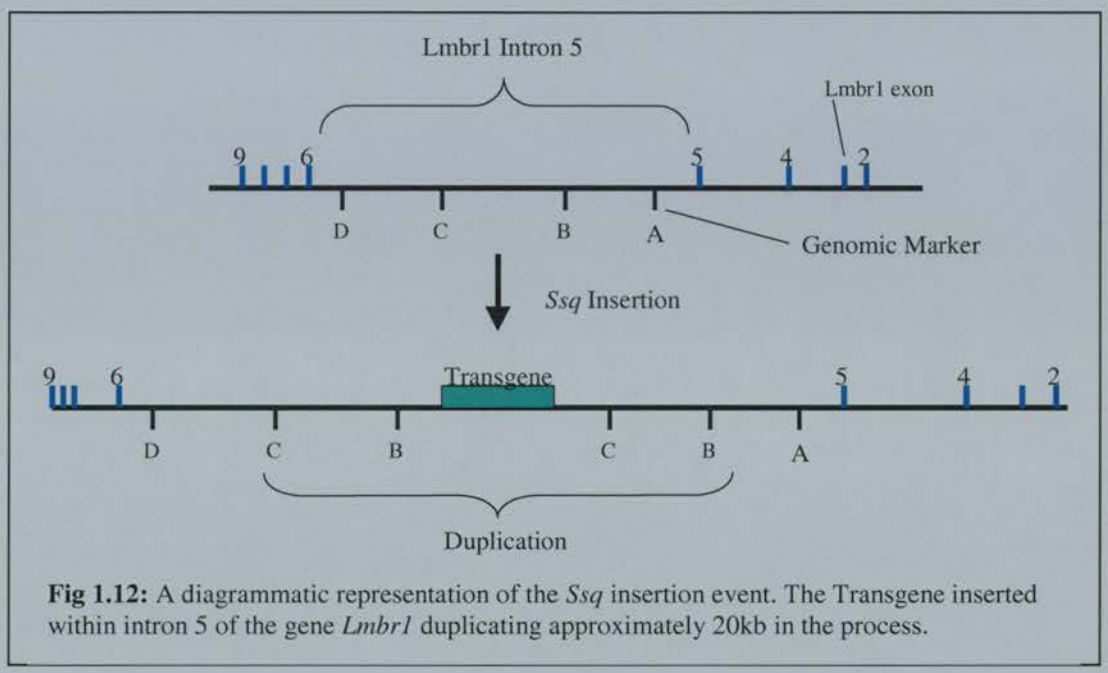


Fig 1.12: A diagrammatic representation of the *Ssq* insertion event. The Transgene inserted within intron 5 of the gene *Lmbr1* duplicating approximately 20kb in the process.

1.10.2 The Lmbr1 gene

Lmbr1 (Limb region 1) is located approximately 800 kb downstream from *Shh* and encodes a 490 amino acid protein (LMBR1L) and possibly a 32 amino acid peptide by alternative splicing (LMBR1S) (Clark et al., 2000). LMBR1L is hydrophobic and predicted to contain nine transmembrane domains, and exhibits over 95% identity to the human gene *C7orf2*, henceforth referred to as *LMBR1*.

Dr. Laura Lettice examined the consequence of the Ssq insertion on Lmbr1 transcription; full-length *Lmbr1* transcripts were found in *Ssq* homozygous embryos (Lettice et al., 2002). However, the majority of *Lmbr1* transcripts in *Ssq* embryos show some premature termination. The 5' end of the transcript is produced at wild type levels, but transcription of exons 3' of the *Ssq* insertion occurs at 10-30% of wild type levels. The *Ssq* transgene is believed to be the cause of the pre-mature truncation between exons 5 and 6 of *Lmbr1 Ssq* animals.

Due to the perturbation of the *Lmbr1* transcript in *Ssq* animals *Lmbr1* was originally thought to play a role in generating the *Ssq* phenotype. A scenario was envisaged whereby the premature truncation of the *Lmbr1* gene acted to create a dominant negative form of LMBR1, that either "switched on", or failed to repress, *Shh* expression in the anterior limb bud. Work from other groups on human and mouse limb mutations that mapped close to *Lmbr1* (outlined below) seemed to support the notion that *Lmbr1* had a role in limb development, probably by acting to regulate *Shh* expression in the limb. However despite the efforts of several groups including our own, significant expression of *Lmbr1* cannot be detected in a developmentally relevant pattern in the mouse limb bud by *in situ* hybridisation (unpublished observations). RT-PCR analysis does detect low levels of *Lmbr1* transcript in all embryo/adult tissues examined, and *Lmbr1* ESTs have been found *in silico* in many embryonic and adult EST libraries. Suggesting that *Lmbr1* is expressed at low levels ubiquitously in the mouse.

1.10.3 The Hx mutation

Lmbr1 has also been implicated in the mouse limb mutation *Hx*, another member of the *hemimelia-luxate* group of mutations, mentioned above. Genetic mapping by Clark *et al.*(Clark et al., 2000) has identified a critical region for *Hx* that includes *Lmbr1* and another gene *Lmbr2* (*RNF32* in humans). Due to the similarity in phenotype of *Hx* and *Ssq,* and the fact they map to similar genomic locations we suspect that *Hx* is allelic to *Ssq*. The open reading frame of both *Lmbr1* and *Lmbr2* were shown to be intact in *Hx* mice. However, the expression of both long and short *Lmbr1* transcripts was demonstrated, by Northern blot analysis, to be absent from *Hx* homozygous hind limb buds at E11.5, and by E12.0 absent from both fore and hind limb buds. The levels of *Lmbr1* transcription returned to normal levels by E12.5. The levels of *Lmbr1* transcription were only mildly down-regulated in *Hx* heterozygotes at E12.0. *Hx* homozygotes and *Hx* heterozygotes have an indistinguishable phenotype at both pre-natal and post-natal stages. *Lmbr2* transcripts were not detected by

Clark *et al.* in either wild type or *Hx* limb buds. Due to the down-regulation of *Lmbr1* transcripts in *Hx* limbs at E11.5 (the same embryonic stage at which morphological defects occur in *Hx* limb buds) Clark *et al.* postulated that the *Hx* mutation disrupted a regulatory element of *Lmbr1*. Leading to the down regulation of *Lmbr1* in the limb bud, which in turn resulted in the *Hx* limb phenotype.

1.10.4 Lmbr1 knockout

In an effort to establish if *Lmbr1* had a role in limb development Clark *et al.* generated a loss of function mutation in *Lmbr1* using stem cell targeting (Clark et al., 2001). Homologous recombination in ES cells was used to replace a 1.1kb genomic fragment with a PGK*neo* selection cassette. This 1.1 kb fragment included the 5'-most coding exon of the *Lmbr1* gene (exon 1) which was known to encode the first 22 amino acids of both the LMBR1L and LMBR1S proteins. The deleted fragment contains 365bp 5' of exon 1 and 696bp 3', the allele generated by ES cell recombination was termed $Lmbr1^{ATG}$. $Lmbr1^{ATG}$ homozygotes were generated at normal Mendelian ratios by crossing $Lmbr1^{ATG}$ heterozygotes.

To assess if the $Lmbr1^{ATG}$ allele was a true null allele of the *Lmbr1* gene, Northern blots were carried out using adult brain poly(A) RNA from wild type and $Lmbr1^{ATG/ATG}$ mice and a *Lmbr1* cDNA probe. Normal *Lmbr1*, 3 and 5 kb messages containing the first *Lmbr1* exon were greatly reduced in $Lmbr1^{ATG/ATG}$ mice. Long exposures of $Lmbr1^{ATG/ATG}$ Northern blots showed novel multiple *Lmbr1* transcripts in brain RNA. *Lmbr1* transcripts were produced at approximately 7% of wild type levels. Clark *et al.* suggest that the *Lmbr1* transcripts seen may initiate from an alternative promoter within the region or from within the PGK*neo* selection cassette, used to create the $Lmbr1^{ATG}$ allele. They conclude that that the $Lmbr1^{ATG}$ mutation is hypomorphic rather than a null mutation of the *Lmbr1* gene. However we find it difficult to envisage a way in which a functional LMBR1 protein could be produced from the transcripts observed in $Lmbr1^{ATG/ATG}$ mice.

Lmbr1^{ATG} heterozygous mice did not exhibit any digit abnormalities, however homozygotes did display a low incidence of limb abnormalities. Out of 36 mice examined, two displayed loss of a digit on a single limb and syndactyly (joining of the digits), a further mouse displayed syndactyly only. Mice *trans*-heterozygous for the *Lmbr1*^{ATG} allele and the *Hdh*^{dfJ} deletion (Schimenti et al., 2000) (a deletion that removes a large section of mouse chromosome 5 including *Shh* and *Lmbr1*) were also generated and their phenotypes examined. *Lmbr1*^{ATG}/*Hdh*^{dfJ} mice showed consistently more severe limb phenotypes, for example 55% of all *Lmbr1*^{ATG}/*Hdh*^{dfJ} limbs had fewer than five digits, compared to <1% of *Lmbr1*$^{ATG/ATG}$ limbs. In the light of *Lmbr1*$^{ATG/ATG}$ and *Lmbr1*^{ATG}/*Hdh*^{dfJ} phenotypes Clark *et al.* have suggested that *Lmbr1* is essential for the normal development of autopodial limb structures.

1.11 Human Limb mutations mapped to 7q36

1.11.1 Limb Specific Preaxial polydactyly

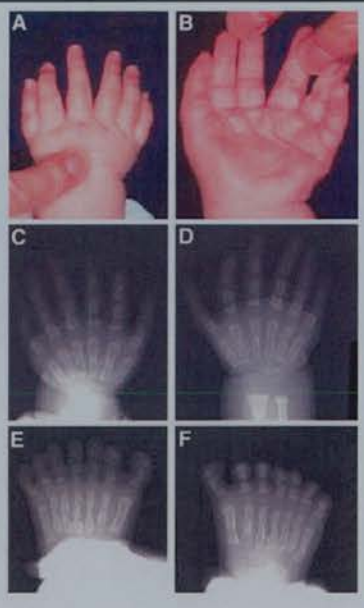


Fig 1.13: An example of human Pre-axial Polydactyly. (Taken from Lettice *et al.*)

Preaxial polydactyly (PPD) is a congenital hand malformation that includes duplicated thumbs, various forms of triphalangeal thumbs and duplications of the index finger (see Fig 1.13). Linkage analysis has located a PPD locus near to the polymorphic marker D7S559 on human chromosome 7q36 (Heutink et al., 1994), and a critical region identified between D7S559 and D7S2423, a distance of approximately 1.9cM (Zguricas et al., 1999). Interestingly, this region is syntenic to the *Lmbr1* region of mouse chromosome 5. This particular form of PPD is limb specific; i.e. it is not associated with any other congenital abnormality, only patient's limbs are

affected. The similarity in phenotype and genomic location between *Ssq/Hx* and limb specific PPD have led us to suggest that *Ssq* and *Hx* are the mouse models for PPD.

A high-resolution genomic map of the 7q36 PPD locus has been created (Heus et al., 1999), and using a combination of exon trapping, cDNA selection and EST mapping 11 transcripts were identified between D7S559 and D7S2423. Further refinement of the PPD critical region with novel polymorphic markers, enabled the location of the PPD mutation to be narrowed down to a 450kb region (Heus et al., 1999). Four transcripts were mapped to this PPD candidate region, *HLXB9,* (a homeobox-containing transcription factor) C7orf3, *RNF32* (both of unknown function) and *LMBR1*. No coding mutations could be found in any of the four genes. Additionally, *LMBR1* transcripts have been amplified by RT-PCR from lymphoblastoid cell lines derived from several PPD families (Lettice et al., 2002). The *LMBR1* transcripts did not exhibit pre-mature truncation, and were identical to wild type controls. This data argues that mutations disrupting *LMBR1* transcription or function are not commonly associated with a PPD phenotype.

1.11.2 A Translocation breakpoint associated with a Japanese PPD patient

Recently a 3-year-old Japanese PPD patient has been identified, and found to carry a *de novo* reciprocal translocation t(5,7)(q11,q36) (Lettice et al., 2002). Fine mapping of the translocation breakpoint, has identified it as lying within intron 5 of *LMBR1*. The presence of the translocation was predicted to form a truncated *Lmbr1* transcript similar to those formed in *Ssq* mice.

1.11.3 Acheiorpodia

Acheiropodia is an autosomal recessive developmental disorder that results in bilateral congenital amputations of the upper and lower limbs, plus aplasia of the hands and feet (see Fig 1.14) (Ianakiev et al., 2001). No other systemic manifestations have been reported, thus acheiropodia is a limb specific disorder. Analysis of five families with acheiropodia by Ianakiev *et al.*(Ianakiev et al., 2001) identified a critical region for the

Fig 1.14: A +B An Acheiropodia affected individual, **C + D** leg x-rays from the same individual (taken frome Ianakiev *et al* 2001).

The *sasquatch* mouse: an enhanced limb mutation on 7q36 between polymorphic markers D7S3037 and D7S3036, a distance of <0.5Mb. Three genes were contained within this critical region *C7orf3, RNF32* and *LMBR1*. Analysis of the *LMBR1* gene in acheiropodia individuals revealed that they contained a deletion, the boundaries being 1.2-2.5 kb 5' and 2.7-3.5 kb 3' of *LMBR1* exon 4. Thus, the acheiropodia deletion leads to the production of an *LMBR1* transcript lacking exon 4, and introduces a frameshift that leads to a premature stop codon in exon 6. Due to these results Ianakiev *et al*. postulated that *Lmbr1* plays a vital role in limb development, and is required for limb outgrowth.

1.12 Lmbr1 a gene vital for limb development?

All the limb mutations outlined above contain *Lmbr1/LMBR1* within their critical regions, and many have been shown to directly disrupt the *Lmbr1/LMBR1* transcript (see Fig 1.15). Thus initially it was suspected that *Lmbr1* might play a role in limb development,

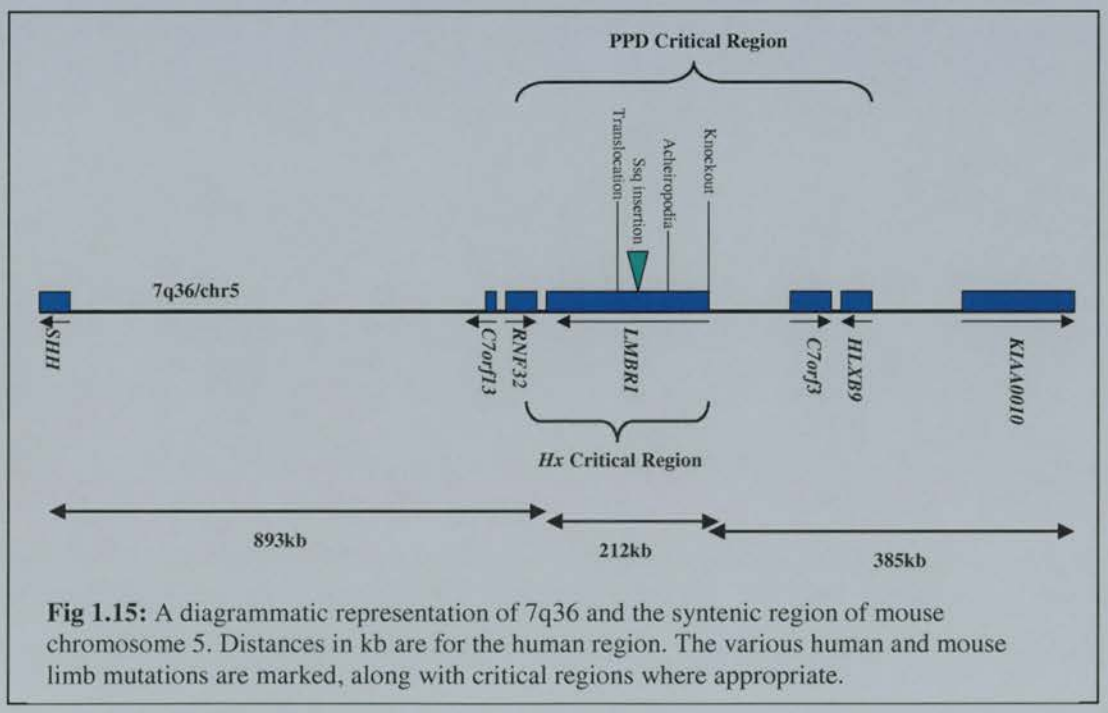


Fig 1.15: A diagrammatic representation of 7q36 and the syntenic region of mouse chromosome 5. Distances in kb are for the human region. The various human and mouse limb mutations are marked, along with critical regions where appropriate.

probably as a regulator of *Shh* expression in the limb. It was suggested that the truncated transcripts in the *Ssq* mouse, and Japanese translocation could result in a dominant positive form of LMBR1, resulting in the up-regulation of *Shh* in the anterior limb bud, and therefore extra digits. The formation of null or partially null alleles in acheiropodia and $Lmbr1^{ATG}$ mice respectively, were envisaged to down-regulate *Lmbr1* and therefore inhibit or reduce *Shh* expression in the limb bud, leading to digit loss and truncations. Although the "*Lmbr1* regulates *Shh*" model was initially attractive, several inconsistencies remained.

Firstly, with the exception of the Japanese translocation most patients that exhibit limb specific PPD do not show any *Lmbr1* structural mutations or alterations in *Lmbr1* transcription. Obviously this is difficult to resolve, as these patients would be expected to exhibit similar dominant positive *Lmbr1* transcripts as the *Ssq* mouse and *Japanese* PPD patient. Secondly, the *Hx* mouse despite having a PPD-like phenotype, exhibits a complete loss of *Lmbr1* transcription at a critical developmental stage, this is inconsistent with the formation of a dominantly positive form of LMBR1. Thus, it would be expected to have a phenotype more akin to acheiropodia or $Lmbr1^{ATG}/Hdh^{df4J}$ mice, i.e. digit loss. Finally, *Lmbr1* is expressed ubiquitously in all tissues so far examined, but developmental genes tend to show specific expression patterns in the tissues they help to form. Thus, if *Lmbr1* were acting to regulate *Shh* expression in the limb, we would expect *Lmbr1* to exhibit a limb specific expression pattern. Could these inconsistencies with the "*Lmbr1* regulates *Shh*" model be pointing to an alternative mechanism that could explain the limb phenotypes mapped to human 7q36 and mouse chromosome 5? A further aspect of the *Ssq* mutation seemed to suggest that this could be the case.

1.13 The sasquatch Mouse Provides an Alternative Paradigm

As has already been mentioned the *Ssq* line was generated by the random insertion of a transgene that included a *HPAP* reporter (Sharpe et al., 1999). Intriguingly, unlike the other seven lines generated with the same construct, when *Ssq* embryos were assayed for

HPAP activity, they exhibited *HPAP* expression in the limb in addition to the typical rhombomere-4 pattern mediated by *Hoxb1* elements within the transgene (Sharpe et al., 1999). The limb pattern was first detected in the ZPA at E10.5 and it closely parallels that of endogenous *Shh* in the limb. *Ssq* heterozygous animals were also shown to exhibit *HPAP* activity in the anterior region of hindlimb buds, in a spatial and temporal location that mirrors the anterior ectopic *Shh* expression observed in *Ssq* heterozygotes. Homozygous embryos demonstrate anterior *HPAP* staining in fore-limbs as well, consistent with the appearance of ectopic *Shh*. Thus, the HPAP reporter at the *Ssq* insertion site mirrors both normal and ectopic expression of *Shh* in the limb buds of *Ssq* heterozygotes and homozygotes (see Fig 1.16).

Fig 1.16: *HPAP* staining of *Ssq* heterozygote limb buds. *HPAP* expression closely mirrors that of *Shh* in both the posterior and anterior of the limb bud.

These observations strongly suggest that the *HPAP* transgene at the *Ssq* insertion has come under the influence of *cis*-acting gene regulatory elements that drive expression in the limb bud. A possible scenario is that these elements are required for driving *Shh* expression in the limb, and that the *Ssq* insertion event has not only revealed their presence, but also disrupted their activity, such that *Shh* becomes anteriorly expressed in *Ssq* limb buds. This "regulatory hypothesis" is attractive as it does not require the limb mutations mentioned

above to affect the expression of *Lmbr1*. Instead, the various genetic lesions mapped to human 7q36/mouse chr5 could act to disrupt *cis*-acting regulatory elements of *Shh*.

1.14 Thesis Aim

The major aim of this thesis is to further our understanding of limb development, particularly the process by which *Shh* is localised to the ZPA of the limb bud. One strategy with which to approach this aim is to study mutants in which the normal localisation of *Shh* to the ZPA has been disrupted. The *Ssq* mouse is such a mutant, and has already suggested two possible factors that act to position *Shh* to the ZPA, the gene *Lmbr1,* and limb specific regulatory elements of *Shh*. The first section of the thesis resolves which of these factors underlies the *Ssq* mutation and can be summarised as the following question, "Is the *Ssq* phenotype the result of a disruption of *Lmbr1* or *cis*-acting regulatory elements of *Shh*"? The second part of the thesis concerns the identification of candidate sequences responsible for the limb expression of *HPAP* observed in *Ssq* embryos.

Chapter 2

Materials and Methods

2.1 General Methods

2.1.1 Manipulation of nucleic acids

Reagents:
All chemicals were analytical grade and were supplied by Sigma, Promega, Gibco BRL, BDH, Fisher Scientific, Flowgen, and Roche. Nucleic acid manipulations were done in 1.5ml centrifuge tubes unless otherwise stated. General solutions were prepared by HGU technical staff and autoclaved and stored at room temperature.

Tris.HCl
Tris base (tris[hydroxymethyl]aminomethane) was dissolved in sterile water. HCl was used to adjust the pH to the required value.

EDTA
EDTA (ethyldiaminetetra-acetic acid di-sodium salt) was dissolved in sterile distilled water. The solution was adjusted to pH 8.0 by adding solid NaOH.

TE buffer
10mM Tris.HCl (pH 7.5); 1mM EDTA.

TBE buffer, 20X stock

Tris base	216g
Boric acid	110g
0.5M EDTA	80mM

Distilled water was added to a final volume of 1 litre. Stock was diluted to 1X with distilled water.

TAE 50X stock

Tris base	242g
Glacial acetic acid	57.1ml
0.5M EDTA	100ml

Distilled water was added to a final volume of 1 litre. Stock was diluted to 1X with distilled water.

Electrophoresis:
Agarose gel loading buffer
Loading buffer was prepared as a 10X stock and stored at room temperature.

	Final concentration
Ficoll	20%
Orange G (Sigma)	1%
EDTA	100mM

Made to required volume with distilled water.

For preparing gels, the required amount of agarose (High pure, BioGene) was dissolved in either 1X TBE or TAE by heating. Molten agarose was cooled and ethidium bromide added to a final concentration of 10µg/100ml agarose. DNA size markers were run alongside experimental samples to estimate the size and amount of DNA in the samples. The size marker used routinely was 1 Kb DNA ladder (Gibco BRL Cat. No. 15615-016). Samples were run in 1X loading buffer. Electrophoresis was used when DNA fragments of a particular size needed to be purified. To do this, samples were electrophoresed using low melting point agarose and the DNA was recovered using a gel extraction kit (Qiagen), according to the manufacturer's instructions.

Determining the concentration of DNA samples:
DNA concentrations were determined in one of two ways: by agarose gel electrophoresis or by measuring the absorbency in a spectrophotometer at 260nm (A_{260}). To determine the concentration by electrophoresis, several different volumes of the DNA (e.g., 1,3 and 5µl) were run alongside standard amounts of DNA (usually DNA size markers). An estimate of the concentration was made by visual comparison of the samples with the known amounts of DNA under UV illumination.

To determine the concentration with a spectrophotometer, the DNA sample was diluted 1:100 with dH_2O. The spectrophotometer was calibrated using a water only blank sample. The samples were placed in clean cuvettes and the absorbance at 260nm (A_{260}) was measured. The concentration of the original sample in mg/ml was calculated as follows.

Concentration (mg/ml) = A260 x 100(dilution factor) x 50.

Restriction enzyme digestion:
All restriction enzymes were purchased from Roche unless otherwise stated. Digests were ideally carried out in a large volume (usually 50-100µl) to minimise effects of evaporation. Enzyme was added at a concentration of 5-10u/µg DNA, depending on the duration of the digests. For overnight digests, 5u/µl were used. The manufacturer's guidelines were followed to determine the correct buffer and incubation temperature for each enzyme.

Klenow enzyme was used at a concentration of 1u/µg DNA to end fill digested ends of DNA following digestion with certain restriction enzymes. The reaction was done at room temperature for 15 minutes and it was stopped by heat inactivation at 75°C for 10 minutes, as recommended by the manufacturer.

Removal of buffer salts:
In cases where multiple restriction digestion steps were done and the enzymes required different buffers, digests were done separately. Buffer salts were removed between digests by drop dialysis and new buffers were added. A nitro-cellulose filter (Millipore) was placed in a petri dish containing filter sterilised dH_2O. The digestion reaction was carefully placed on the filter (a maximum of 50µl per filter) and salts were left to diffuse through the filter for at least 1 hour at room temperature. Following dialysis, the liquid was removed to a fresh 1.5ml centrifuge tube and the appropriate enzyme and buffer were added for the subsequent reaction.

As an alternative to using a filter, salts were removed using either a PCR cleanup kit or a nucleotide removal kit (Qiagen).

Ligations:
In order to maximise the ligation efficiency, ligations were set up with the following vector: insert molar ratios: 1:3, 1:5, and vector only (to control for re-ligation of the vector). The amount of vector in each case was 25ng, and the total reaction volume was 10µl. Ligations were done at 16°C overnight using DNA Ligase and ligation buffer (Boehringer Mannheim).

2.1.2 Microbiology

Aseptic technique was observed for all steps involving the growth of bacterial cells (setting up cultures, pouring agar plates, selecting single colonies, and storing bacterial stocks). Liquid cultures were grown at 37°C with vigorous shaking and dry cultures were grown on inverted agar plates at 37°C. HGU technical staff prepared all bacterial growth media stocks as follows:

L Broth
Amounts per litre:

Tryptone	10.0g
Yeast extract	5.0g
NaCl	10.0g
Glucose	1.0g

Production of electrocompetent cells:
A single colony of XL1 Blue *E.coli* cells from an agar plate was used to inoculate approximately 10ml L broth for overnight growth. The culture was used to inoculate 2 X 400ml fresh L broth the following morning. Cells were grown to an absorbance at 600nm (A_{600}) of 0.15-1.0. Flasks were chilled on ice for 15-30 minutes and cells were centrifuged at 4°C, 4000g for 15 minutes. Pellets were

re-suspended in 800ml ice cold sterile 10% glycerol. The centrifugation step was repeated and cells were re-suspended in 400ml 10% cold glycerol. Following a further centrifugation step, cells were re-suspended in 20ml 10% glycerol. Cells were centrifuged once again and re-suspended in 2-3ml 10% glycerol. The final concentration of cells was approximately $3x10^{10}$ cells/ml. Aliquots of cells were frozen in liquid nitrogen and stored at –70°C.

Transformations:
Competent cells were transformed with DNA by electroporation. For each ligation, 1µl DNA was placed in an ice cold centrifuge tube. Electrocompetent cells were thawed on ice and 50µl were added to each DNA sample. The mixture was transferred to ice cold electroporation cuvettes and allowed to sit for 1 minute on ice. Cells were electroporated using a BioRad electroporator set at 25µF, 2.5kV, and 200Ω. 1ml L broth/ Mg^{2+} was added to the cells immediately following electroporation and cells allowed to recover for 1hour by shaking at 37°C prior to plating on selective medium.

Isolation of plasmid DNA:
Plasmid DNA was prepared using commercially available kits (Qiagen or ABI). For the extraction of small amounts of plasmid DNA (minipreps), a single colony was used to inoculate approximately 10ml L broth with antibiotic selection (ampicillin) for overnight growth. The following morning, plasmid DNA was extracted using the kit according to the manufacturer's instructions. Plasmid DNA was eluted in 30 or 50µl elution buffer. The DNA concentration was determined either by agarose gel electrophoresis or by spectrophotometer.

2.1.3 Polymerase Chain Reaction (PCR)

dNTPs
dNTPs were purchased as stocks of 100mM. Working stocks of 10mM were made by mixing 10µl of each of the dNTPs (dATP, dCTP, dGTP, dTTP) with 60µl dH_2O to a final volume of 100µl. Stocks were stored at –20°C. dNTPs were used in PCRs at a final concentration of 0.2mM.

Primers
Primers were purchased from MWG Biotech as lyophilised desalted compounds. Stocks were made up to 100µM using sterile dH_2O. Primers were used in PCRs at a final concentration of 1µM (1:100 dilution).

Taq polymerase, PCR buffer, and Mg^{2+}
These reagents were purchased (Applied Biosystems). Ampli*Taq* (5u/µl) was used at 0.2µl per 25µl reaction. PCR buffer was a 10X stock and therefore diluted 1:10 for reactions. Mg^{2+} was used at a final concentration of 1.5mM unless otherwise stated. Routine PCRs were done in a MJ Research DNA Engine Tetrad.

PCR amplification programme:
Generally all PCR amplification programmes were identical except for the annealing temperature which was varied according to the primers used:

General PCR amplification programme:

1. 95°C for 3 mins
2. 95°C for 30 secs
3. n°C for 30 secs (annealing temperature)
4. 72°C for 1 min 30 secs
5. Goto step 2 x 30
6. 72°C for 5 mins

2.1.4 Sequencing

Plasmid DNA used for sequencing was prepared using a Qiagen miniprep kit, according to the manufacturer's instructions, and salts were removed using either a PCR cleanup kit or a nucleotide

removal kit column (both supplied by Qiagen). DNA was sequenced using dye-labelled terminators for cycle sequencing. The dye used was either dRhodamine or Big Dye Terminator RR mix. Reagents were thawed on ice and dyes were protected from light as much as possible. Reactions were set up in 0.2ml centrifuge tubes as follows.

200-500ng plasmid DNA in dH_2O	11µl
dRhodamine or Big Dye Terminator RR mix	8µl
primer (3.2 pmoles)	1µl
	20µl

Cycle sequencing was performed using an MJ Research DNA Engine Tetrad. The sequencing program was as follows: [96°C, 30 seconds; 50°C, 15 seconds; 60°C, 4 minutes] for 24 cycles per reaction.
Reactions were ethanol precipitated following transfer to a fresh 1.5ml centrifuge tube containing 50µl ethanol and 2µl sodium acetate, pH 5.2. Reactions were left at room temperature to precipitate for one to eight hours, then centrifuged at 13,000g at 4°C for 30 minutes. Pellets were washed with 200µl 70% ethanol. The supernatant was removed following a second centrifugation step and pellets were allowed to dry at room temperature with the caps off for approximately 20 minutes. Samples were submitted to the sequencing service to be run on an ABI machine.

2.1.5 Animal husbandry

Animals used during transgenic procedures were maintained in a specific pathogen free (SPF) environment, all other animals were maintained in semi-barrier unit. All experiments were carried out under Home Office licence. Wild type animals (CBA, CD1 and C57BL/6) were either bred in-house or obtained from Charles River Laboratories. Breeding animals were maintained on a CBA background. Embryos for all experiments were generated from timed matings, with the morning of vaginal plug detection being counted as embryonic day 0.5 (E0.5).

Genotyping of breeding mice:
Breeding animals were genotyped by PCR amplification of allelic sequences from genomic DNA extracted from tail biopsies. Approximately 1cm tail tissue was removed from the ends of the tails of anaesthetised animals. Tail tips were digested in 1.5ml microfuge tubes at 55°C overnight in 500µl tail tip buffer/ proteinase K. The following day, tubes were vortexed and centrifuged at full speed for 10 minutes. 500µl isopropanol was added to the supernatant and tubes were shaken to mix the solutions. Genomic DNA was spooled out to a fresh centrifuge tube containing 500µl dH_2O using a sterile pipette tip. The solution was pipetted gently to resuspend the DNA. PCR was done using 1µl of DNA (approximately 50ng) and either of the following pairs of primers:

Harvesting of Postimplantation embryos:
Postimplantation embryos were harvested for wholemount *HPAP* and *LacZ* staining. The mothers were sacrificed by cervical dislocation. The abdominal cavity was opened and uteri were removed to petri dishes containing phosphate buffered saline (PBS) (Oxoid, Unipath). Embryos were removed from the uteri and freed of extraembryonic membranes using scissors and forceps. Extraembryonic membranes were retained to genotype embryos by PCR. Embryos were rinsed in fresh PBS prior to subsequent processing.

Genotyping of embryos:
Embryos processed for *HPAP* or *LacZ* staining were genotyped using DNA derived from extra-embryonic membranes. In both cases, DNA was extracted and amplified using the same procedure as described for tail biopsies.

Tail tip/Embryo sac lysis buffer:

	Final concentration
Tris.HCl pH8.0	100mM
EDTA	50mM
NaCl	100mM

SDS 1%
Proteinase K (stock 10mg/ml) (Sigma) was added to a final concentration of 0.2mg/ml.

2.2 Chapter 3 Methods

2.2.1 PCR Genotyping assays

Shh alleles:
The general PCR amplification programme was used with an annealing temperature of 59°C. The primer sequences were:
P1: 5' GAC CAT GTC TGC ACA CTT AGG TTC C 3'
P2: 5' GAA GGC CAG GAG GAG AAG GCT CAC 3'
P3: 5' CTG TGC TCG ACG TTG TCA CTG 3'
P4: 5' GAT CCC CTC AGA AGA ACT CGT 3'

Ssq allele:
The general PCR amplification programme was used with an annealing temperature of 53°C. The primer sequences were:
P5: 5' CTC TGT TTC CTT TTC CTC TAT C 3'
p6: 5' GTA TGG GAT TAA TTA AAT CTT GTG TC 3'

2.2.2 Animal Husbandry

The strain of the original Shh^{null} heterozygous female was 129/Sv. All further matings were carried out on a CBA background.

2.3 Chapter 4 Methods

2.3.1 HPAP staining

Protocol:
Embryos were harvested and genotyped as mentioned above, and then fixed in 4% paraformaldehyde (PFA) for 10-20mins and then washed twice in PBS (salt solution) to remove excess PFA. Embryos were then heated in PBS at 65°C for 40mins, and left to cool for 20 mins. *HPAP* staining solution was then added, and the embryos were left at room temperature in the dark until the stain developed. The staining reaction was stopped with 50mM EDTA pH 5.0, and the embryos were re-fixed with 4% PFA overnight.

Solutions:
4% PFA: 20g of paraformaldehyde (Sigma) dissolved in 500mls PBS at 65°C overnight. Aliquots were stored at -20°C

HPAP staining solution:
100mM Tris pH 8.5
100mM NaCl
50mM MgCl2
1mg/ml NBT (Nitroblue tetrazoluim chloride) (Boehringer Mannheim)
0.1mg/ml BCIP (5Bromo-4-chloro indoylphosphate) (Boehringer Mannheim)

2.3.2 Microscopy

Dissections of postimplantation embryos were done with the aid of a Zeiss Stemi SV11 dissecting microscope. Photographs of wholemount embryos were taken using a Photometrics ICX205 digital colour CCD camera and a Leica M2F2III stereoscopic microscope with brightfield illumination. Images were captured using IP Lab software.

2.4 Chapter 5 Methods

2.4.1 Cosmid sequencing

Generation of plasmid DNA for sequencing was carried out using a Beckman Coulter Biomek 2000 Laboratory Automation Workstation. Sequencing reactions were then implemented as described above, using the Biomek 2000 Laboratory Automation Workstation and run on an ABI 3700

2.4.2 Bioinfomatics Programmes

The programmes used to generate the bioinformatic data in chapter 5 are listed below along with, their versions and functions:

BLAST version 2.1.3 Altschul *et al.* 1997 (sequence alignment)
ClustalW version 1.74 Thompson *et al.* 1994 (phylogentic alignment)
Consed version 11.0 Gordon *et al* 1998 (sequence contig building)
EMBOSS version (sequence manipulation)
Mega2 version 2.1 Kumar et al. 2000 (phylogentic alignment)
RepeatMasker version 07/07/2001 Smit, unpublished (comparative sequence analysis)
Wise2 version 2-1-22c Birney, unpublished (sequence annotation)
VISTA version (comparative sequence analysis)
PIPMaker version (comparative sequence analysis)

All programmes were used with their default settings, except when using PIPMaker to compare human to fish sequences, where the chaining and high-sensitivity options were used.

2.4.3 Comparative Sequence Analysis

Sequences to be compared were annotated using BLAST and Wise2. They were then edited using EMBOSS, and compared using VISTA and PIPMaker.

2.4.3 Phylogenetic Analysis

Alignments were created using Clustalw. The output of the alignment was then converted to Mega2 format using the ForCon program. Mega2 was then used to construct and bootstrap the phylogeny. Minimum evolution with complete deletion was used to construct the phylogeny. Bootstrapping was carried out with 1000 repetitions with a random seed.

2.5 Chapter 6 Methods

2.5.1 Production of transgenic animals

Preparation of recombinant DNA for microinjection:
Plasmid DNA was digested (10µg DNA in a 100µl reaction) with enzymes to release the transgene. For each digest, a sample of DNA was run on a 1% agarose gel in TBE alongside an equivalent amount of undigested DNA to ensure the DNA was completely digested and products were the predicted size. Once the transgene had been released, the remainder of the reaction was run on a 1% agarose gel in TAE buffer. The transgene was then extracted from the gel using a DNA gel extraction kit (Qiagen).
The DNA was purified and concentrated using microcon 30 columns (Amicon) according to the manufacturer's instructions. Briefly, the DNA was passed through a microcon column and washed three times using 0.1mM EDTA/1mM TRIS pH7.4. DNA was eluted in 10µl 0.1mM EDTA/1mM TRIS pH7.4. The eluant was diluted 1:10 in transgenic buffer (0.1mM EDTA/10mM TRIS pH7.4).

The DNA concentration was determined by electrophoresis. DNA was stored at –20°C until the day of microinjection. For microinjection, DNA was diluted in transgenic buffer to a final concentration on 2ng/μl and spun through a Spinex 0.22μm column (Costar).

Microinjection of recombinant DNA into fertilised eggs:
Embryos used for microinjection were derived from [CBA x C57BL/6] F1 matings. Females were superovulated by intraperitoneal injection of 10 units of PMS (Intervet) at noon on the first day, followed by 10 units of human chorionic gonadotrophin (hCG) (Intervet) at 2pm two days later. They were then mated to F1 males. Plugged females were sacrificed the following morning (E0.5) and oviducts were removed and rinsed in warm saline. Oviducts were placed in H6 media and fertilised eggs were removed by tearing the swollen ampullae with fine forceps. Several drops of hyaluronidase (1mg/ml PBS) (Sigma) were added to the media for approximately three minutes to remove the cumulus cells from the outside of the eggs. Fertilised eggs were identified by the presence of two large pronuclei and often polar bodies. Eggs awaiting microinjection were kept at 37°C, 5% CO_2 in drops of T6 media under paraffin oil in sterile petri dishes. Eggs were microinjected in batches of 30-50 in a large drop of H6 media under oil in a glass dimple slide (the injection dish). During microinjection, embryos were picked up and held on the holding pipette using mouth pipette tubing. The microinjection needle on the automatic injector was set such that a constant stream of DNA left the pipette. Once the (usually male) pronucleus was in focus, the injection needle was brought into the same plane of focus and used to pierce the *zona pellucida* and egg cell membrane, before entering the pronucleus. The needle was held until the pronucleus swelled slightly, and then removed quickly and cleanly. Following microinjection, eggs were returned to drops of T6 media under paraffin oil and kept at 37°C, 5% CO_2 overnight. For each experiment, five to ten uninjected embryos were retained and cultured overnight to control for problems with media, oil, and incubation conditions.

Oviducal transfers:
Microinjected embryos were screened the following morning. Only embryos that had developed to the two-cell stage were transferred into recipient females. Recipients were anaesthetised with an intraperitoneal injection of 0.4-0.6 ml anaesthetic [0.75ml of Hypnorm (Fentanyl 0.315mg/ml / fluanisone10mg/ml): 4.5ml water: 0.2ml Hypnovel (Midazolom 10mg/2ml)], and the ovarian fat pad isolated and held outside of the body cavity. The bursa surrounding the oviduct was broken carefully to expose the infundibulum of the oviduct. The end of the pulled pipette containing the embryos was inserted into the infundibulum and embryos were delivered by mouth pipette. The successful delivery of the embryos was confirmed when bubbles that had been placed at either end of the embryos in the pipette were seen in the swollen ampulla. Ten to 20 embryos were transferred to each side of the recipient. The skin was closed using Dieffenbachs bulldog clips (Holborn surgical and medical instruments, ltd.) and the animal was allowed to recover in a warmed post-operative cage with solid drink pouches (BS&S). Embryos were harvested at appropriate developmental stages.

LacZ PCR Genotyping Assay:
The general PCR amplification programme was used with an annealing temperature of 60°C. The primer sequences were:
P7: 5' GCG ACT TCC AGT TCA ACA TC 3'
P8: 5' GAT GAG TTT GGA CAA ACC AC 3'
P9: 5' TTA CGT CCA TCG TGG ACA GC 3'
P10: 5' TGG GCT GGG TGT TAG TCT TA 3'

Equipment:
Embryos were transferred between dishes using pulled pasteur pipettes and mouth pipette tubing. A Zeiss Stemi SV11 dissecting microscope was used to aid in manipulating embryos in sterile glass staining blocks and petri dishes.

Microinjection apparatus:
The microinjection apparatus consisted of a Zeiss Axiovert 100 microscope with a variable temperature stage set to 37°C, and mounted holders for the microinjection and holding pipettes (Narashige, Eppendorf). The equipment rested on an anti-vibration table (Carl Zeiss). Embryo holding pipettes were either made in-house using a microforge (MF90) (Narashige) and capillaries (size GC100 T15, Harvard Apparatus) or purchased (Eppendorf). Microinjection pippetes were either

made in-house using an automated pipette puller (Sutter model p87, Harvard Apparatus) and capillaries (size GC100 TF 10, Harvard Apparatus) or purchased (Eppendorf). Microinjection needles were filled with DNA at a concentration of 2ng/μl using pipette filler tips (Eppendorf). DNA was microinjected with the aid of an automated microinjector (Narashige IM 300) to allow a constant flow of DNA at variable pressure.

Solutions:
H6 media was used for handling embryos outside of the incubator. T6 media was used for growing embryos in the incubator (5% CO_2, 37°C). Media was either purchased (Sigma) or prepared in-house by Transgenic Unit staff using embryo culture grade reagents.

Compound	T6 (g/100ml)	H6 (g/100ml)
NaCl	0.472	0.472
KCl	0.011	0.011
NaH_2PO_4	0.0047	0.0047
Sodium pyruvate	0.003	0.003
Glucose	0.1	0.1
$MgCl_26H_2O$	0.01	0.01
$CaCl_22H_2O$	0.026	0.026
Penicillin	0.006	0.006
Streptomycin	0.005	0.005
$NaHCO_3$	0.2106	0.2106
Sodium lactate	0.34ml	0.34ml
Hepes		0.50
Phenol red	0.001	0.001
BSA	0.4	0.4

Acid Tyrode's solution was used to remove the *zonae pellucidae* from eight-cell embryos prior to aggregation. The working aliquot was kept at 4°C. The solution was warmed to 37°C for use.

Compound	g/litre
NaCl	8.0
KCl	0.2
$CaCl_2$	0.2
$MgCl_2$ $6H_2O$	0.1
NaH_2 PO_4H_2O	0.05
Glucose	1.0
$NaHC0_3$	1.0

2.5.2 LacZ Staining

Staining Protocol:
Embryos were dissected in PBS and fixed in 4% PFA at 4°C for at least 1 hour. Following fixation, embryos were first rinsed in PBS and then washed in detergent wash three times for at least 20 minutes each time. Tissues were stained overnight in Xgal staining solution at 37°C in the dark. The following morning, the staining solution was removed and tissues were rinsed in detergent wash then re-fixed in 4% PFA at 4°C overnight. The fix was removed and tissues were washed in PBS.

Solutions:
Detergent Wash:

Compound	Per 500ml	Final concentration
1M phosphate buffer, pH 7.3	48.5ml	0.1M
1M $MgCl_2$ in phosphate buffer	1.0ml	2mM
5% sodium deoxycholate in phosphate buffer	10ml	0.1%
10% Nonidet P-40 in phosphate buffer	1ml	0.02%
1% BSA in phosphate buffer	2.5ml	0.05%

Xgal stain solution:

Compound	Per 250ml	Final concentration
Detergent wash	236ml	
NaCl, 5.3% stock in phosphate buffer	4ml	0.085%
$K_3Fe(CN)_6$, 250mM in dH_2O	5ml	5mM
$K_4Fe(CN)_6$, 250mM in dH_2O	5ml	5mM
Xgal substrate, 100mg/ml in DMF	750µl	0.3mg/ml

Chapter 3

Cis-trans test

3.1 Introduction

The sasquatch (*Ssq*) mouse mutation is known to stimulate ectopic anterior expression of *Shh* in the developing limb bud, which results in a pre-axial polydactyly phenotype (PPD) (Sharpe et al., 1999). However the molecular mechanism disrupted by the *Ssq* mutation is unknown. This chapter of work uses mouse genetics to reveal the mechanism and consequently provide some insights into how normal *Shh* expression is regulated during limb development.

Clues as to how the *Ssq* mutational event disrupted normal limb development were present from the initial characterisation of the mutant (Sharpe et al., 1999). The *Ssq* HPAP transgene had previously been shown to express in a pattern identical to that of *Shh* and both loci were also known to be linked (Sharpe et al., 1999). Together these two observations suggested that *cis*-acting regulatory elements of *Shh* could be involved. It was conceived that the insertion of the *Ssq* transgene near, or within a limb regulatory element of *Shh,* could disrupt the element's activity resulting in erroneous *Shh* expression in the anterior of the limb. This disrupted element could then also be responsible for driving the HPAP activity observed within the limbs of *Ssq* mice (i.e. the *Ssq* transgene was acting as an enhancer trap). Although linked the distance between the *Ssq* insertion and *Shh* was approx. 1 Mb, a considerable distance for a *cis*-acting regulatory element to act over. Thus if this "regulatory hypothesis" were correct it would require a re-evaluation of the process of gene regulation, and raise many interesting possibilities and questions.

An alternative mechanism involved the *Ssq* insertion acting in a more conventional manner by disrupting the activity of *Lmbr1*. It was proposed that *Lmbr1* could be involved in regulating Shh expression in the limb, perhaps by normally stimulating expression in the posterior domain. Insertion of the *Ssq* transgene could result in creating a dominantly active form of Lmbr1 switching on Shh expression in the anterior of the limb bud. The HPAP activity could be subsequently explained by elements controlling Lmbr1 or an unrelated gene

in the *Ssq* genomic region rather than Shh. This hypothesis was supported by the fact that the *Ssq* insertion site was known to reside within a Lmbr1 intron and cause premature truncation in the majority of Lmbr1 transcripts (Lettice et al., 2002). Premature truncation could conceivably create a constitutive active form of the LMBR1, switching on Shh in the anterior limb.

Further evidence in support of a role for *Lmbr1* in limb development came from various human and mouse mutations, which mapped closely to genomic regions that corresponded with the *Ssq* insertion site. Some of these mutations were known to disrupt the *Lmbr1* transcript. Human Acheiropodia (loss of limbs) deleted *Lmbr1* exon 4 (Ianakiev et al., 2001). A human Japanese translocation mapped to *Lmbr1* intron 5 resulted in a truncated *Lmbr1* transcript, and a PPD phenotype (Lettice et al., 2002). The mouse Hx limb mutation (exhibits ectopic pre-axial digits) showed down-regulation of the *Lmbr1* transcript at E11.5-E12.0 (Clark et al., 2000), and replacement of the first exon of *Lmbr1* with PGKneo resulted in a mild loss of digits (Clark et al., 2001).

However, there were inconsistencies that threw doubt on the speculation that *Lmbr1* had a role in limb development. *Lmbr1* did not have a mouse expression pattern consistent with a role in limb development or the HPAP activity observed in *Ssq* animals. Additionally, no structural mutations within the gene could be found in many familial cases of human PPD (Heus et al., 1999). Thus a paradox existed whereby disruption of *Lmbr1* could explain some but not all of the mutations, and the reason for the *Ssq* limb HPAP activity could not be easily attributed to regulatory elements of *Lmbr1*. To ascertain which of the "regulatory" or "*Lmbr1* regulates *Shh*" hypotheses were correct a mouse genetic assay was developed. The test relied on exploiting the assumption that mouse regulatory elements act in *cis* on their target gene, i.e. only on an allele they are physically linked to, whereas genes act via proteins and can therefore act in *trans* (see Fig. 3.1). Thus, if a method can establish how the *Ssq* insertion acts to promote ectopic *Shh* gene expression (in *cis* or *trans*) it can determine if the *Ssq* mutation has disrupted the activity of a long-range regulatory element or a gene.

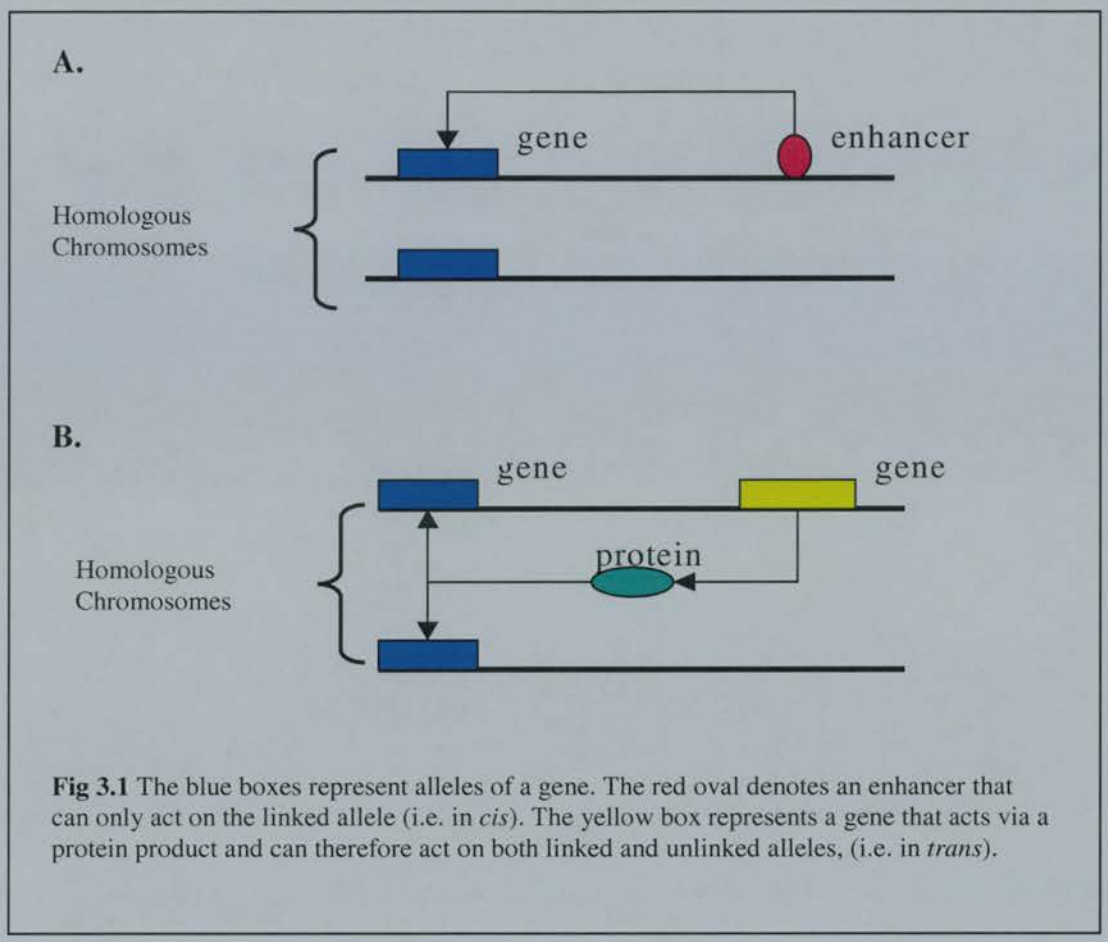


Fig 3.1 The blue boxes represent alleles of a gene. The red oval denotes an enhancer that can only act on the linked allele (i.e. in *cis*). The yellow box represents a gene that acts via a protein product and can therefore act on both linked and unlinked alleles, (i.e. in *trans*).

A "*cis-trans* test" was devised for the *Ssq* mutation using the Shh^{null} allele. It was concluded that if the Shh^{null} allele could be physically linked to the *Ssq* allele within a mouse there could be two possible phenotypes. Acting in *trans* the *Ssq* allele would stimulate ectopic expression of both wild type and null forms of *Shh* in the anterior of the limb bud, resulting in a heterozygous *Ssq* phenotype. Acting in *cis* the *Ssq* allele could only stimulate ectopic anterior expression of non-functional SHH protein from the linked Shh^{null} allele. Non-functional SHH protein would be incapable of patterning the extra digits in the anterior limb bud, resulting in a wild type limb (see fig 3.2). This chapter concerns the creation and analysis of the mice necessary to resolve the *cis-trans* test.

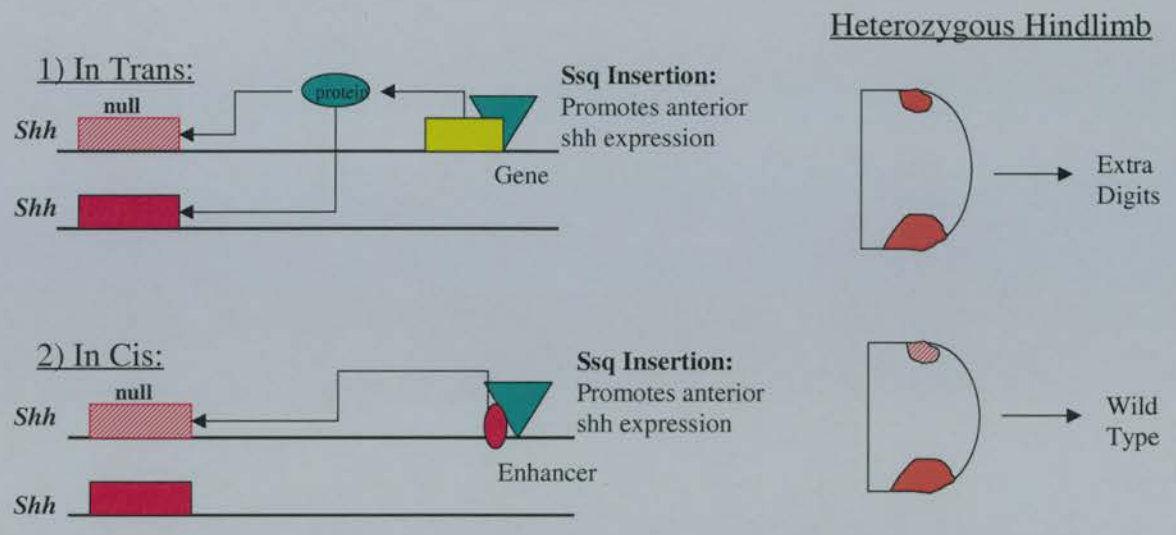


Fig. 3.2: Section 1 describes the outcome if the *Ssq* mutation acts through a gene in trans. The yellow box represents a gene, stimulated by the *Ssq* insertion (green triangle) to produce its protein product in the anterior of the limb bud. Thus switching on expression of both normal and null forms of the *Shh* gene in the anterior domain, resulting in a PPD (extra digits) Phenotype. Section 2 shows the outcome if the *Ssq* mutation acts through a regulatory element of *Shh*. The enhancer (red oval) is stimulated by the Ssq insertion to promote *Shh* expression in the anterior of the limb bud. Because it only acts in *cis* only the null-allele is expressed, resulting in non-functional SHH in the anterior of the limb which cannot pattern ectopic digits resulting in a wild type phenotype.

3.2 Results

3.2.1 PCR Genotyping assays

Crucial to performing the *cis-trans* test was the ability to track the relevant alleles

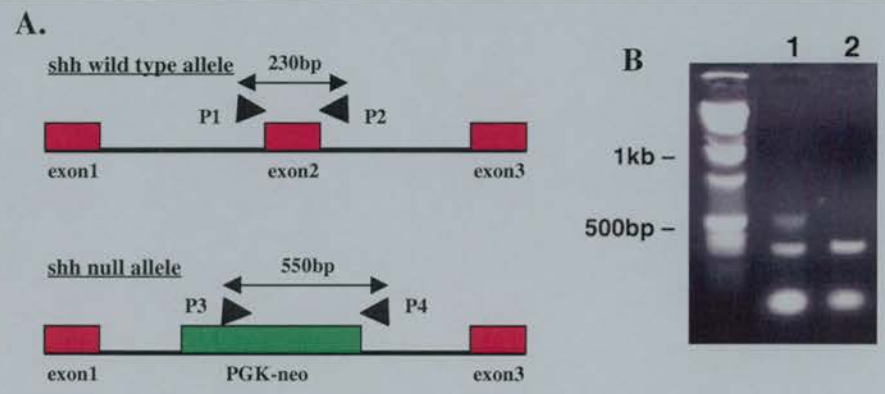


Fig 3.3 Shh allele PCRs: Section A shows a cartoon of the wild type and Shh^{null} alleles. The black triangles represent PCR primer binding sites. Primers P1 and P2 amplify a 230 bp region from exon 2 of the *Shh* gene. Primers P3 and P4 amplify a 550 bp section of the PGK-neo gene used to create the Shh^{null} allele. **Section B** shows PCR results from Shh^{null} heterozygous mice (lane 1) and wild type mice (lane 2).

(*Shh*^{null} and *Ssq*) through several generations of mouse crosses. This could not be done by simple phenotype analysis as wild type, *Shh*^{null} heterozyotes and possibly the mice at the end of the test all had a wild type phenotype.Fortunately, a PCR based assay already existed for detection of the *Shh*^{null} and *Shh* alleles (Chiang et al., 1996). The assay consisted of two sets of primers; one amplified a 550bp band from the neo^{r} gene contained within the *Shh*^{null} allele. The other a 230bp fragment from exon 2 of the *Shh* wildtype allele (removed in the knockout) (see Fig 3.3).

A PCR assay for the *Ssq* allele had to be developed. Dr. Laura Lettice kindly provided sequences from the *Ssq* transgene and the genomic region adjacent to the *Ssq* insertion. Using this sequence several sets of primers were designed and tested for specificity to the *Ssq* allele. A pair of primers, one internal the other external to the Transgene were chosen to track the *Ssq* allele throughout the *cis-trans* test (see fig 3.4).

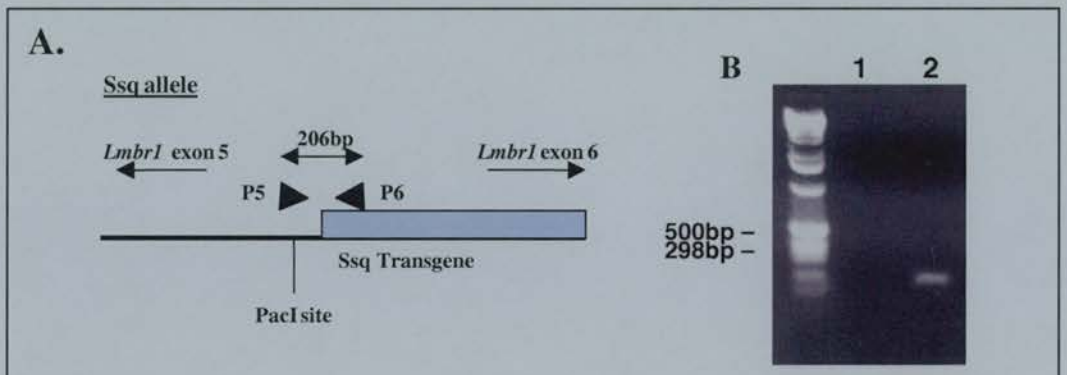


Fig 3.4 The *Ssq* allele PCR: Section A shows a cartoon depicting the binding sites for the two primers (P5 and P6) used to track the *Ssq* allele through the *cis-trans* test. The primers are shown as black triangles and amplify a 206 bp product. Primer P6 binds within the transgene therefore in wild type mice no product is produced using P5 and P6. Section B shows PCR results from wild type mice (lane 1) and *Ssq* heterozygous (lane 2) using primers P5 and P6.

3.2.2 The cis-trans test

The *cis-trans* test relied on physically linking the alleles *Ssq* and *Shh*^{null} on mouse chromosome 5. Due to the distance between these two alleles approx. 1Mb it was conceived

that the simplest way of doing this was via a chromosomal recombination event. With this aim a mouse breeding regime was devised (see Fig. 3.5).

Initially a female mouse heterozygous for the Shh^{null} allele was crossed to a homozygous *Ssq* male to produce the F1 generation. The F1 generation (n = 14) carrying the two alleles on opposite homologous chromosomes (in *trans*) showed complete penetrance of the *Ssq* phenotype. Five F1 males and four F1 females were then mated to wild type mice to generate 446 offspring; the G2 generation. Each G2 mouse was phenotyped and assayed by PCR as described in section 3.2 for the presence of the Shh^{null} and *Ssq* alleles. Examples of the genotyping PCR's are shown in Fig 3.5.

Fig 3.6: The first mouse generated carrying the Shh^{null} and *Ssq* alleles in *cis*. The mouse displayed complete suppression of the *Ssq* phenotype.

The G2 generation consisted of 220 *Ssq* heterozygotes and 218 Shh^{null} heterozygotes, an approximately 1:1 ratio as would be expected from Mendalian genetics. The remaining 8 G2 mice were recombinants of the two alleles representing a recombination frequency of 1.8%.

Three were uninformative as they contained only wild-type alleles. The remaining five recombinants (3 males and 2 females) carried both *Ssq* and Shh^{null} alleles in *cis* (i.e. linked on the same chromosome). Phenotype analysis of these 5 recombinants showed no pre-axial polydactyly or other detectable limb phenotypes (see Fig. 3.6).

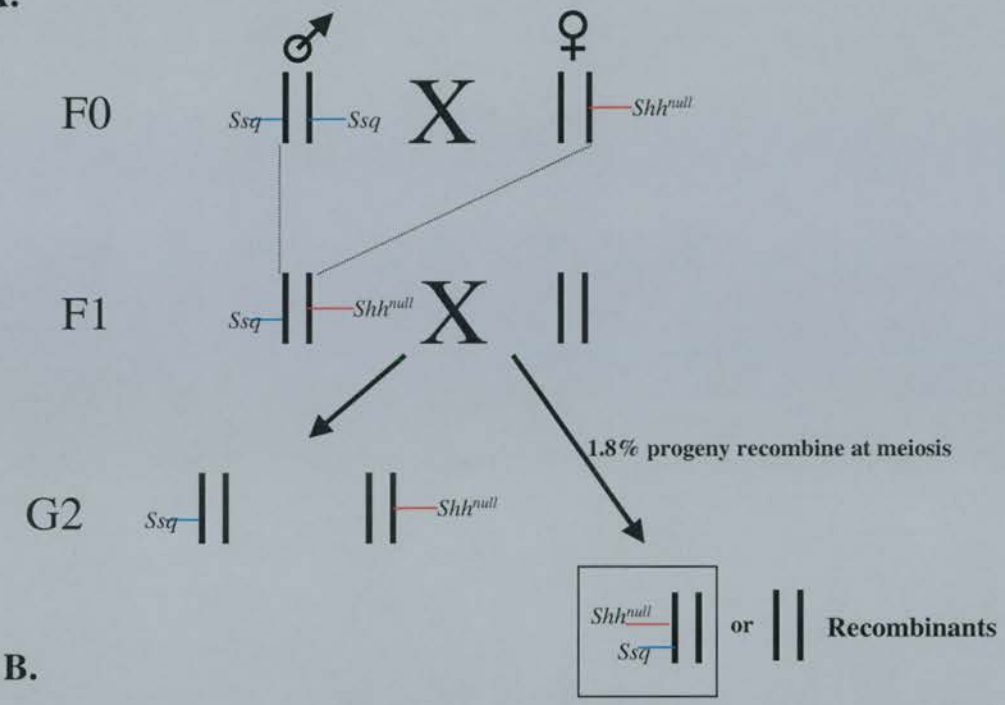
3.2.3 Issues of Penetrance

The *Ssq* heterozygous phenotype is not fully penetrant, the degree depending on the genetic background. During the *cis-trans* test (section 3.3) the *Ssq* heterozygous phenotype was 93% penetrant. 7% of the *Ssq*/+ mice generated showed no detectable limb phenotype.

Statistically it was highly unlikely that suppression of the *Ssq* phenotype observed in the 5 mice carrying the recombinant chromosome was not due the fact they were carrying the *Ssq/Shh*^{null} recombinant chromosome, but due to non-penetrance. The probability of any 5 mice selected at random from the *cis-trans* cross carrying one copy of the *Ssq* allele and all being suppressed due to non-penetrance is, 1.68×10^{-6}.

To alleviate any further concerns regarding penetrance, 2 further crosses were set up (see Fig.3.7). Firstly 2 non-penetrant G2 *Ssq* heterozygous males were crossed to wild type females, and were shown to transmit the phenotype; 16 with pre-axial polydactyly out of 36 offspring. Secondly two males carrying the recombinant chromosome 5 (*Ssq/Shh*^{null}) were bred with wild-type females. All 49 mice generated displayed a wild type phenotype, including the 25 carrying the recombinant chromosome.

A.



B.

Genotype:	***shh*^{null}/+**	***Ssq*/+**	**+/+**	***shh*^{null}Ssq/+**
Number:	**218**	**220**	**3**	**5**
	wild type	**PPD**	**wild type**	**wild type**

C.

Fig 3.5: Section A shows a diagrammatic representation of the *cis-trans* test. The horizontal black lines represent mouse chromosome 5, the *Ssq* and *Shh*^{null} alleles are marked. Non-marked alleles are wild type. Section B records the results of the *cis-trans* test, showing the number of each mouse genotype produced and its phenotype. The genotype carrying the *Ssq* and *Shh*^{null} alleles in *cis* is highlighted in red. Section C is a photograph of a sample of the genotyping PCR's carried out during the *cis-trans* test. The white asterix highlights the PCR results from a recombinant mouse generated during the *cis-trans* test which carried the *Ssq* and *Shh*^{null} alleles in *cis*.

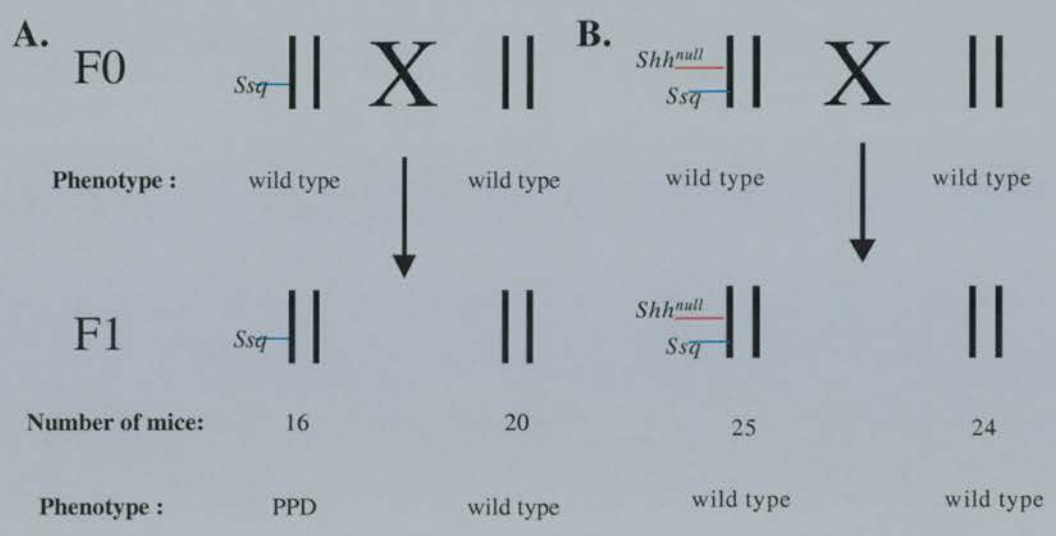


Fig 3.7: A diagrammatic representation of the two mouse crosses used to establish if the phenotypic suppression seen in the *cis-trans* test is due to linkage of the Shh^{null} and *Ssq* alleles or non-penetrance. The black vertical lines represent mouse chromosome 5, the Shh^{null} and *Ssq* alleles are marked appropriately. **Section A** shows the cross used to establish that suppressed *Ssq* heterozygotes can transmit the PPD phenotype. **Section B**, reveals the cross used to demonstrate that mice carrying the Shh^{null} and *Ssq* alleles in *cis*, transmit a suppression of the PPD phenotype to their offspring.

3.2.4 Reversion Cross

The informative recombinants generated during the *cis-trans* test showed suppression of the *Ssq* heterozygous phenotype, a result that could also be explained by the *Ssq* allele having lost its ability to promote anterior *Shh* expression (i.e. become non-functional). Although the fact that five separate informative recombinants were generated would tend to argue against this, (it seemed highly unlikely *Ssq* function could be lost in all five). It was deemed prudent to check that the *Ssq* allele on the recombinant chromosome 5 retained the capacity to produce a PPD phenotype.

With this aim in mind a further mouse cross (the "Reversion Cross") was devised and implemented (see Fig 3.8) to generate a recombination event between the Shh^{null} and *Ssq* alleles on the recombinant chromosome 5. This event would separate the alleles and reverse the original *cis-trans* cross. The *Ssq* heterozygote offspring generated from this reversion

cross could then be phenotyped and assessed for PPD. The presence of PPD would confirm the fact that the *Ssq* allele had maintained its capacity to form ectopic digits and was suppressed due to the presence of the Shh^{null} allele in *cis*, not due to loss of function from the *Ssq* allele. From the previous cis-trans cross in section 3.3 the expected frequency of recombination was 1.8%.

Five males carrying the recombinant chromosome were crossed to wild-type females producing 438 F1 offspring, all of which were wild type. 246 of these mice were genotyped as described in section 3.2, and no recombination events were observed; all the mice had a wild type or Shh^{null} *Ssq*/+ genotype. The chance of not observing a recombination event in the 246 mice genotyped in the reversion cross is 0.012 (see below), assuming a recombination frequency of 1.8%. Due to the fact that no recombination events were observed it was impossible to confirm that the *Ssq* allele had retained its activity at this stage.

Poisson Distribution:

$Pr(n) = e^{-\mu} \times \mu^n/n!$ Where Pr(n) = probability of n events
μ = expected frequency
n = number of events

Thus: $Pr(0) = 2.718^{-4.4} \times 4.4^0/0! = 0.012 \times 1/1 = 0.012$

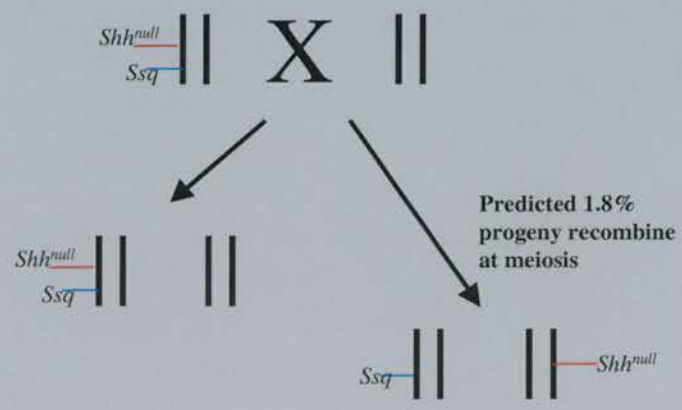


Fig 3.8: is a diagrammatic representation of the reversion cross, an attempt to separate the Shh^{null} and *Ssq* alleles linked by the *cis-trans* cross (section 3.3). Based on the *cis-trans* cross, 1.8% of the progeny from the reversion cross were predicted to recombine at meiosis between the Shh^{null} and *Ssq* alleles thus unlinking them. Unfortunately no such recombination event was observed either phenotypically or genotypically.

3.3 Discussion

3.3.1 Cis-Trans Test

All five of the informative recombinant mice generated completely suppressed the *Ssq* phenotype. The data in section 3.4 confirmed that non-penetrance of *Ssq* was not the cause of this suppression as mice bred from non-penetrant males exhibited PPD but those from the recombinants did not. Therefore the *Ssq* mutation can only act in *cis*, i.e. it can only promote ectopic expression from the *Shh* allele it is physically linked to. Thus the regulatory hypothesis is correct, and *Ssq* is actually a regulatory allele of *Shh*, henceforth refered to as Shh^{Ssq}.

This conclusion raises some interesting problems. Firstly, Shh^{Ssq} is a gain of function mutation and this is hard to reconcile with a simple regulatory model such as a single limb enhancer element. Disruption of a tissue specific enhancer would be envisaged to remove expression of a gene from a particular tissue, but in the Shh^{Ssq} limb *Shh* gains an expression field in the anterior of the limb bud. Thus it is reasonable to assume that a regulatory model explaining the Shh^{Ssq} mutation must include a repressor element (see Fig. 3.9). This repressor would normally inhibit the activity of a limb enhancer in the anterior limb bud, but the repressor rather than the enhancer is disrupted by the *Ssq* insertion. Thus the un-repressed limb enhancer in Shh^{Ssq} embryos is free to promote expression of both *Shh* and *HPAP* in the normal posterior and ectopic anterior domains of the limb bud. At this stage the location of these *Shh* limb regulatory elements is unknown. Though it is probably safe to assume that at least the repressor part of the regulator is some where close to the *Ssq* insertion site as it is thought to be disrupted by the mutation. However, the enhancer which may be a separate element, could be some distance away; if it can act over a large distance on *Shh* it could be acting long-range on HPAP. Results of a hunt for the limb regulatory elements of *Shh* are recorded in chapters 5 and 6.

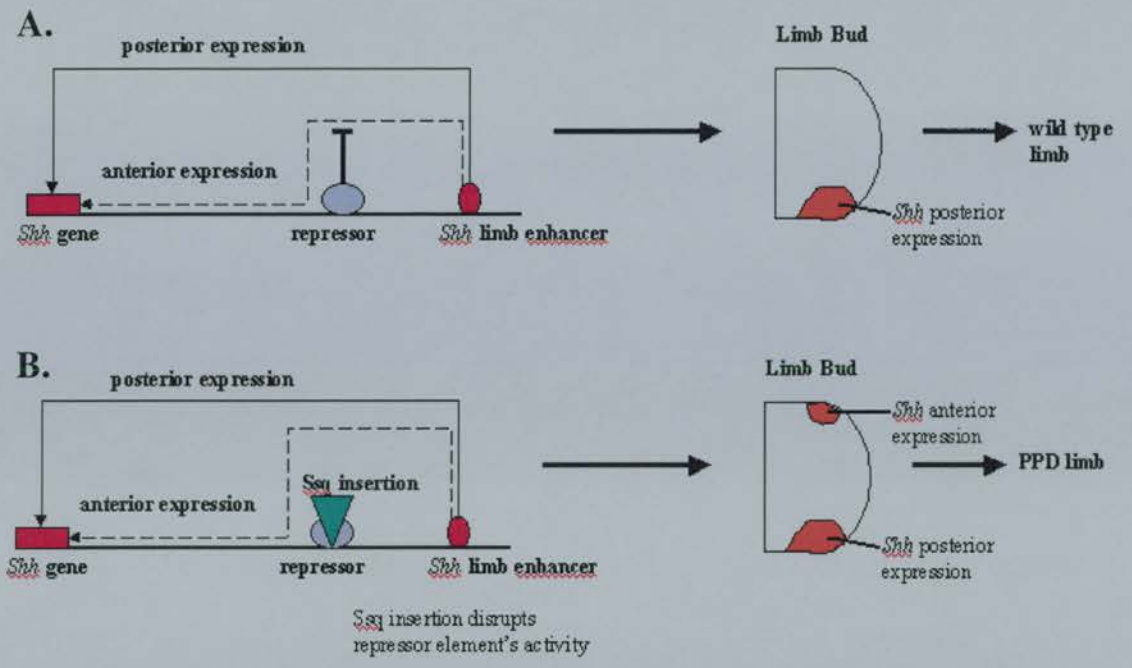


Fig 3.9 The limb enhancer/repressor model. **A** shows how the model is proposed to work in wild type limbs. The *Shh* limb enhancer drives expression in both anterior and posterior domains in the limb bud, but the repressor inhibits the anterior expression. Thus *Shh* is only expressed in the posterior domain. **B** demonstrates the model working in a sasquatch mouse hindlimb. The insertion removes the activity of the repressor, allowing the enhancer to drive expression of *Shh* in both anterior and posterior domains, resulting in a PPD phenotype.

3.3.2 *Lmbr1*

The *cis-trans* test suggests that the disruption to the *Lmbr1* transcript observed in Shh^{Ssq} mice is co-incidental to the mutation and that *Lmbr1* does not have a role during limb development. This is consistent with work on familial human PPD patients that exhibited no structural mutations within the *Lmbr1* gene (Heus et al., 1999). However other mouse and human limb mutations which map to the *Lmbr1* region need a more detailed explanation with regards to the *cis-trans* test, and the involvement *of Shh* limb regulators. The mouse *Hx* mutation which is probably allelic to Shh^{Ssq} does show down-regulation of the *Lmbr1* transcript but no structural mutations can be found (Clark et al., 2000), so as for Shh^{Ssq} perhaps disruption of *Lmbr1* is co-incidental to the mutation. The Japanese human translocation, which displays a PPD phenotype (Lettice et al., 2002), does truncate the

Lmbr1 transcript but it also "disconnects" a large part of the *Lmbr1* genomic region from *Shh*, perhaps separating a repressor element from an enhancer sequence; thus stimulating *Shh* expression in the anterior of the limb bud and creating a PPD phenotype. This theory suggests a possible location for an enhancer sequence, between the site of translocation and *Shh*, as the enhancer would have to remain linked to *Shh* in the absence of the repressor.

Two final recessive mutations to consider are human Acheiropodia (Ianakiev et al., 2001) and the mouse *Lmbr1* knockout generated by Clark *et al.*(Clark et al., 2001). Initially both these limb mutations seemed to strongly suggest a role for *Lmbr1* in limb development, as truncated or non-functional protein products were predicted from the mutational events. However with regards to the *cis-trans* data, both mutations remove genomic DNA surrounding exons of *Lmbr1*, which may contain enhancer sequences involved in limb regulation of *Shh*. Removal of such sequences would down-regulate *Shh* specifically in the limb (more severely in Acheiropodia), but only for individuals homozygous for the mutation. Fitting the characteristics of both mutations perfectly and suggesting second and third possible locations for enhancer sequences near to *Lmbr1* exons 1 and 4.

When combined with these human and mouse mutations the *cis-trans* data seems to suggest a complex regulatory model for *Shh* in the limb, consisting of at least four elements scattered throughout the *Lmbr1* gene (see Fig 3.10). Three or more enhancer sequences could be acting in concert to drive *Shh* expression in the anterior and posterior limb bud. The repressor element then acts to "fine tune" the enhancer by repressing it in the anterior limb bud, resulting in a wild type expression pattern.

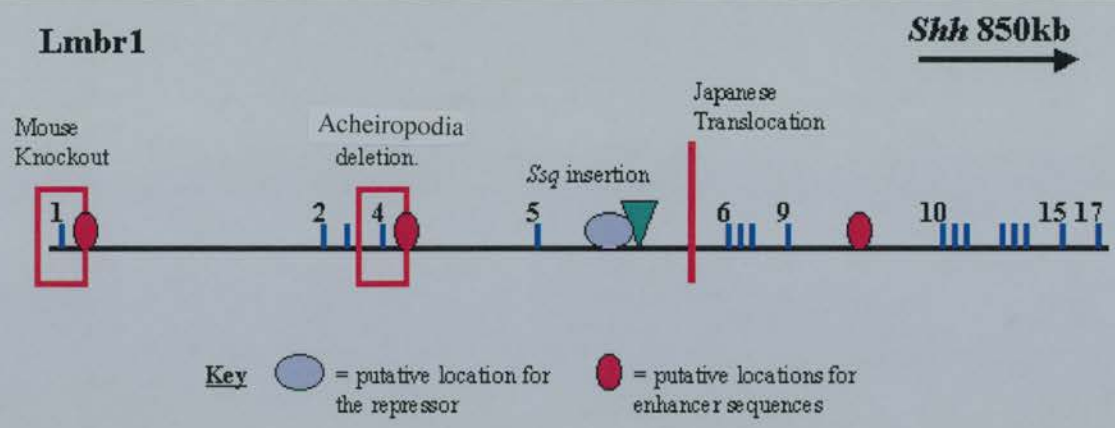


Fig 3.10 Combining the *cis-trans* data with human and mouse mutations in the *Lmbr1* region suggests a complex limb regulatory model. Mutations that result in the loss of digits suggest the removal of enhancer sequences that drive *Shh* expression in the limb. Mutations that result in PPD suggest disruption of a repressor that down-regulates anterior *Shh* limb expression. Thus the mutations suggest possible locations for enhancer and repressor sequences. However the mutations are scattered throughout *Lmbr1* requiring there to be multiple enhancer sequences that act in concert to drive expression of *Shh* in the limb.

3.3.3 Reversion Cross

The *cis-trans* test has provided a clear insight into the mechanism of *Shh* regulation in the developing limb bud. However further genetic attempts to confirm the *cis-trans* results by reversing the recombination failed. 250 mice were genotyped, 4-5 were expected to recombine between the Shh^{null} and Shh^{Ssq} alleles but none of the 250 exhibited a recombination event. The reason for this inhibition of recombination remains unclear and is particularly puzzling due to the fact that the two loci recombined at a consistent rate in the *cis-trans* test. Currently there seems to be no satisfactory explanation for this data other than chance. It is possible that not enough mice were looked at during the cross and we were simply "unlucky" in not seeing a recombination event. However due to financial and time constraints it was decided not to pursue the reversion cross further. Further data in results chapter 4 confirms the presence of an active *Ssq* transgene in the recombinant mice, and the remainder of this thesis adds further evidence supporting the validity of the *cis-trans* result.

Chapter 4

HPAP activity in recombinant mice

4.1 Introduction

The sasquatch (*Ssq*) mutation was generated via a random insertion of a *Hoxb1/human placental alkaline phosphatase* (HPAP) transgene into intron 5 of the gene *Lmbr1* (Sharpe et al., 1999). *Hoxb1* elements within this transgene were known to drive expression of HPAP in rhombomere-4, but in *Ssq* embryos HPAP activity was also seen in the limb bud from E10.5 (Sharpe et al., 1999). The HPAP limb pattern was unique to this insertion site and closely paralleled the expression pattern of *Shh*. HPAP staining was detected in the ZPA, the normal site of *Shh* expression, but from E12.5 HPAP was also present in the anterior region of limb buds showing anterior ectopic expression of *Shh*. *Ssq* heterozygous embryos showed ectopic anterior HPAP expression in hind limbs only, whereas *Ssq* homozygotes exhibited anterior ectopic HPAP activity in fore and hind limbs. As mentioned in Chapters 1 and 3 the HPAP limb activity observed in *Ssq* embryos was presumed to be the result of the action of *cis*-acting *Shh* limb regulatory elements driving expression of the HPAP reporter. The *cis-trans* test in chapter 3 linked the *Ssq* and Shh^{null} alleles, and confirmed that the *Ssq* mutation acted by disrupting these *Shh* limb regulatory elements to promote ectopic anterior *Shh* expression in the developing limb bud.

The HPAP expression data in recombinant embryos could provide several important pieces of information regarding the regulation of *Shh* in the limb. Firstly such data would contribute to the validity of the *cis-trans* test, by demonstrating that recombinant embryos still contained *Ssq* alleles capable of driving appropriate HPAP activity within the limb. Confirming that the Ssq allele had not itself been lost or rendered incapable of promoting gene expression in the anterior hind limbs of $Shh^{null,Ssq}$ embryos.

Secondly homozygous recombinant embryos (embryos with two copies of the recombinant chromosome, $Shh^{null,Ssq}/Shh^{null,Ssq}$) could be used to study the activity of the *Ssq* regulatory elements in the absence of functional SHH. When combined with data from other groups regarding the expression of genes thought to be important in controlling *Shh* limb

expression such as *dHand* and *Gli3*, it was thought that some important insights into *Shh* regulation in the limb would be revealed. The collection and analysis of HPAP reporter activity in recombinant embryos is recorded in this chapter.

4.2 Results

4.2.1 HPAP activity within recombinant embryos

Recombinant mice generated during the *cis-trans* test carried the *Ssq* and Shh^{null} alleles in *cis* and exhibited complete suppression of the *Ssq* phenotype. Despite being phenotypically normal recombinant mice were presumed to still carry an active HPAP transgene under the influence of the *Shh* limb regulatory elements discussed in chapter 3. Therefore the HPAP transgene was predicted to have retained a *Ssq* heterozygote limb expression pattern in recombinant embryos for both posterior and anterior domains (see Fig 4.1).

To ascertain the presence of HPAP activity in the recombinant mice, males

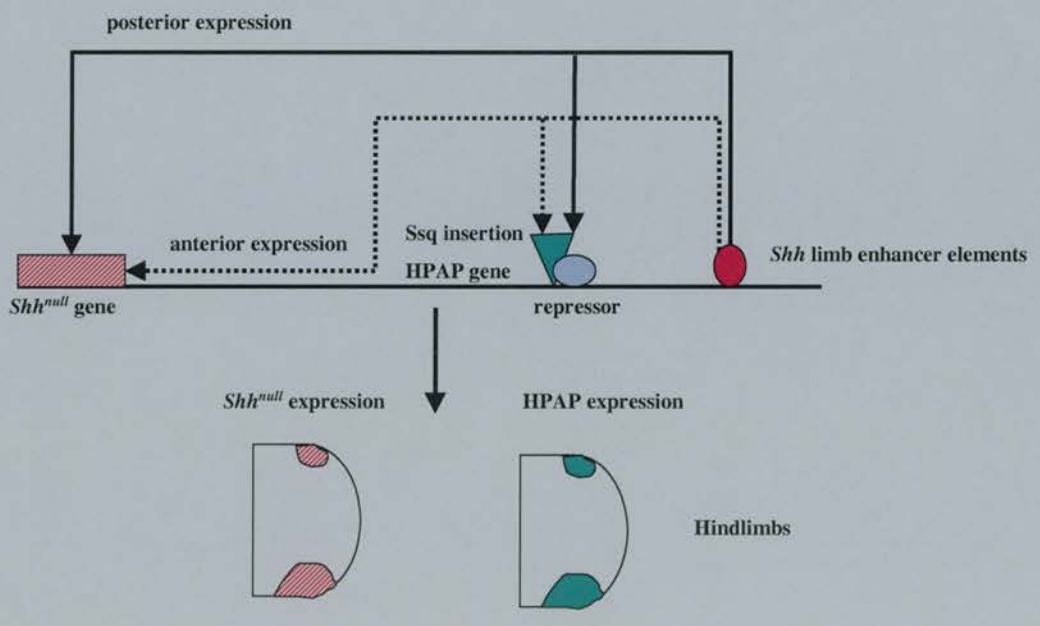


Fig. 4.1 Predicted HPAP activity in recombinant embryos. The *Ssq* HPAP gene in recombinant embryos is believed to be under the influence of the same regulatory elements that are driving expression of the Shh^{null} allele. Thus it was predicted that HPAP would be expressed in the same limb pattern as the Shh^{null} allele.

heterozygous for the recombinant chromosome were crossed to CBA wild type females. *Ssq* heterozygous males were also crossed to CBA wild type females to act as a positive control for the HPAP staining. Embryos were harvested from these crosses at E10.5, E11.5 and E12.5, and then stained for HPAP activity as described in Chapter 2.

Figure 4.2 shows heterozygous *Ssq* (*Ssq/+)* and recombinant ($Shh^{null,Ssq}$/+) embryos from all 3 stages stained with HPAP. As expected all 3 stages of the *Ssq/+* embryos exhibited HPAP activity in the ZPA of the fore and hind limbs. At E12.5 HPAP activity was also observed in the anterior of the limb bud, at the tip of the ectopic tissue that forms the ectopic digits seen in *Ssq* animals. A band of HPAP staining was also seen in rhombomere-4 at E10.5, fading at E11.5 and absent by E12.5. $Shh^{null,Ssq}$/+ embryos showed similar HPAP expression patterns in the ZPA and rhomobmere-4, but the anterior HPAP staining at E12.5 was absent along with the ectopic anterior tissue in which it is normally situated. This is more clearly seen in Figs. 4.2 panel B and C.

Figure 4.2 panel B shows whole wild type, *Ssq/+* and $Shh^{null,Ssq}$/+ E12.5 embryos in more detail. The HPAP staining in the *Ssq/+* and $Shh^{null,Ssq}$/+ embryos is pronounced but the anterior ectopic tissue and HPAP staining is absent from $Shh^{null,Ssq}$/+ hind limbs. This is shown more clearly in Figure 4.2 panel C, which shows the dissected limbs from E12.5 wild type, *Ssq/+* and $Shh^{null,Ssq}$/+ embryos.

From the above results it is clear that the recombinant embryos do carry a copy of the *Ssq* transgene, as HPAP activity is evident in the ZPA of both fore and hind limbs. However none of the E12.5 recombinant embryos (n=23) exhibited HPAP staining or ectopic tissue in the anterior limb bud. A further investigation into this result is contained in the next section.

4.2.2 *Recombinant cross to* luxate

Recombinant embryos carrying the *Ssq* and Shh^{null} alleles in *cis* failed to express HPAP in the anterior of the limb at E12.5. The reason for this was unclear; one possible

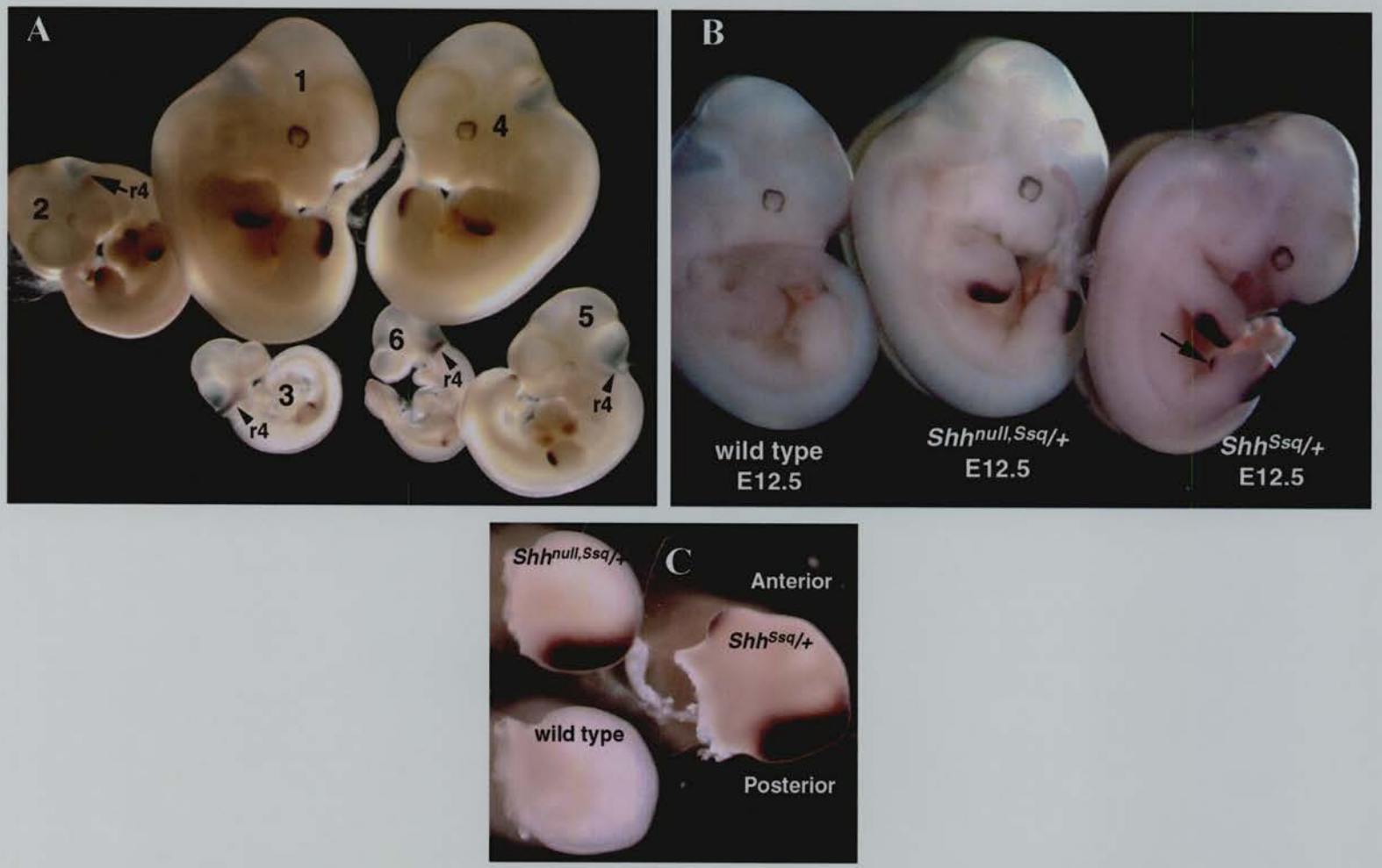


Fig 4.2: A: Embryos 1, 2 and 3 are $Shh^{Ssq}/+$ stages E12.5, E11.5 and E10.5 respectively. Embryos 4, 5 and 6 are $Shh^{null,Ssq}/+$ stages E12.5, E11.5 and E10.5 respectively. The black arrows highlight *HPAP* staining in rhombomere 4. **B:** From left to right, wild type, $Shh^{null,Ssq}/+$ and $Shh^{Ssq}/+$ E12.5 embryos. The arrow highlights the anterior limb *HPAP* stain **C:** The dissected hind-limbs from E12.5 wild type, $Shh^{null,Ssq}/+$ and $Shh^{Ssq}/+$ embryos, notice the lack of anterior *HPAP* activity in $Shh^{null,Ssq}/+$ limbs

theory was that the *Ssq* transgene had somehow lost the ability to promote gene expression in the anterior limb bud of recombinant embryos. This was considered unlikely due to the fact that the recombinant embryos assayed for HPAP came from three of the five original recombinants generated during the *cis-trans* test. It was difficult to envisage a process in which the *Ssq* transgene had lost its ability to promote anterior HPAP expression in all three parent mice.

A more feasible alternative theory was based on the fact that recombinant embryos fail to express functional *Shh* in the anterior of the limb bud, which is believed to be essential for the creation of the anterior ectopic limb tissue seen in Shh^{Ssq}/+ mice. Without functional *Shh* to promote cell proliferation, a large region of the anterior limb bud undergoes apoptosis, probably killing the cells that would normally express HPAP in the anterior limb bud. Consistent with this hypothesis is the fact that anterior HPAP expression pattern in E12.5 Shh^{Ssq}/+ limbs was always observed to lie at the very tip of the ectopic tissue (see Fig 4.2 panel C). Recombinant embryos do not posses this tissue probably due to a lack of anterior functional *Shh*, therefore they do not have the conventional site for anterior HPAP expression in the limb.

To test this "tissue-less" theory, recombinant males were crossed to females heterozygous for the semi-dominant limb mutation *luxate* (*lx*), which displays pre-axial polydactyly. It was hoped that mice heterozygous for the *lx* mutation and carrying a recombinant chromosome would be produced. These *lx* recombinant embryos were predicted to generate ectopic anterior limb tissue in the same manner as *Ssq* heterozygotes, due to the presence of the *lx* loci. Thus it would be possible to ascertain the limb HPAP expression pattern of the recombinant chromosome in the presence of the anterior ectopic limb tissue. According to the "tissue-less" theory when the anterior ectopic limb tissue is restored to the recombinant embryo's limb HPAP would then be able to be expressed in the anterior of the limb bud.

An E12.5 *lx* recombinant embryo is shown in figure 4.3 along with E12.5 *lx* heterozygous and $Shh^{null,Ssq}$/+ embryos. HPAP staining is present as expected in the ZPA of the $Shh^{null,Ssq}$/+ and the *lx* recombinant embryos, but only the embryo with ectopic anterior limb tissue (*lx* recombinant embryo) exhibits the anterior hind limb HPAP expression. Thus the presence of the anterior ectopic limb tissue is required for the expression of the anterior ectopic HPAP expression in the limb bud. The *lx* recombinant embryo demonstrates that the *Ssq* transgene on the recombinant chromosome has retained its ability to promote anterior gene expression in the developing limb.

4.2.3 HPAP activity within homozygous recombinant embryos

The regulatory elements highlighted by the *Ssq* mutation have been suggested to directly promote *Shh* expression within the developing limb. For this statement to be true the *Ssq* regulatory elements must be acting epistatically upstream of *Shh*. Therefore in the absence of active SHH the *Ssq* regulatory elements should still drive expression of HPAP in a *Shh* pattern within the limb bud.

To test the ability of the *Ssq* regulatory elements ability to drive HPAP in the absence of SHH, male and female recombinant ($Shh^{null,Ssq}$/+) mice were crossed to each other and embryos harvested at E11.0. In addition to wild type and $Shh^{null,Ssq}$/+ embryos the cross also produced homozygous recombinants ($Shh^{null,Ssq}/Shh^{null,Ssq}$) that contained two copies of the recombinant chromosome which linked the *Ssq* and Shh^{null} alleles. Thus in these homozygous recombinants it was possible to assay for HPAP activity (i.e. action of the *Ssq* regulatory elements) in the absence of functional SHH. Figure 4.4 shows a homozygous recombinant embryo and one of its wild type and $Shh^{null,Ssq}$/+ littermates. The homozygous recombinant embryo shows severe deformities in the forebrain and midbrain structures and a general reduction in size as expected due to the absence of functional SHH (Chiang et al., 1996). Limb morphology at this stage is delayed relative to its wild type and $Shh^{null,Ssq}$/+

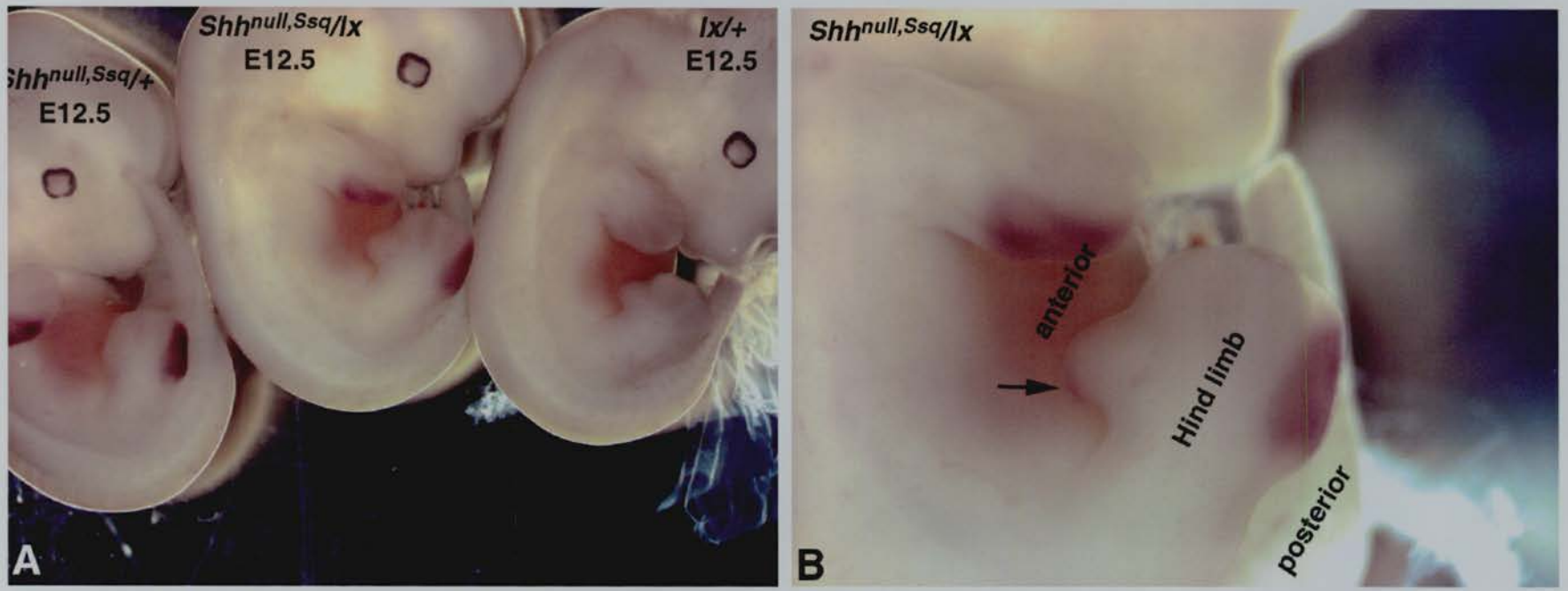


Fig 4.3: A: E12.5 embryos stained for *HPAP* activity, the embryo genotypes from left to right are, $Shh^{null,Ssq}/+$, $Shh^{null,Ssq}/lx$ and $lx/+$. **B:** shows a close up of a hindlimb from the $Shh^{null,Ssq}/lx$ embryo shown in panel A. *HPAP* staining at the tip of the ectopic anterior is highlighted by the black arrow.

Fig 4.4: E11.0 embryos stained for *HPAP*. **A:** A $Shh^{null,Ssq}/Shh^{null,Ssq}$ embryo, exhibiting staining in the ZPA of both fore and hind limbs. However some staining is also observed in the anterior of the limb bud, and in the spinal cord. **B:** A $Shh^{null,Ssq}/+$ embryo exhibiting a similar staining pattern to the $Shh^{null,Ssq}/Shh^{null,Ssq}$ embryo in panel A. **C:** A wild type embryo for comparison.

littermates, but fore and hind limb structures can be seen. Limb HPAP staining is present in the ZPA of fore and hind limbs of the homozygous recombinant. Thus even in the absence of functional SHH the *Ssq* HPAP transgene is still stimulated by the *Ssq* limb regulatory elements highlighted in chapter 3.

HPAP staining of $Shh^{null,Ssq}$/ $Shh^{null,Ssq}$ and $Shh^{null,Ssq}$/+ embryos at E11.0 (n=11) also revealed previously unreported HPAP activity in the spinal cord and anterior forelimbs (see Fig.4.4). Although this staining was weaker than the activity previously observed in the ZPA, it was not thought to be due to endogenous alkaline phosphatase activity, as similar HPAP activity was not apparent in wild type embryos. Thus these two new domains of HPAP expression appeared to be genuine. The reason why they have escaped the attention of previous studies (Sharpe et al., 1999) is possibly due to their specificity of expression, which seems to occur precisely at E11.0, a stage not previously examined.

4.3 Discussion

4.3.1 HPAP in recombinant embryos ($Shh^{null,Ssq}$/+)

The data in section 4.2 clearly demonstrates that mice carrying the *Ssq* and Shh^{null} alleles in *cis* still produce HPAP in the ZPA of the limb, confirming the presence of the *Ssq* allele in recombinant mice. However the full Ssq heterozygous HPAP limb expression pattern was only seen when the recombinant chromosome was crossed onto the *lx* mutation (section 4.3). Thus the anterior HPAP limb expression was always co-incident with the presence of anterior ectopic limb tissue.

The generation of anterior ectopic limb tissue is seen in all polydactylous limb mutants, in most cases due to the expression of *Shh* in the anterior limb bud. Anterior SHH is thought to create an ectopic anterior ZPA that is believed to interfere in two aspects of normal limb development. Firstly, stimulating ectopic cell proliferation to generate the anterior limb tissue that eventually develops into ectopic digits. Secondly the anterior ZPA is

believed to inhibit the apoptosis that normally occurs in the anterior of the limb bud (Sanz-Ezquerro and Tickle, 2000).

$Shh^{null,Ssq}$/+ mice only express from a null *Shh* in the anterior of the limb bud therefore no anterior ZPA is formed, no anterior cell proliferation takes place and anterior apoptosis occurs as normal. We suspect that initially HPAP is expressed along with the Shh^{null} allele in the anterior of the limb bud, but the lack of an anterior ZPA probably results in non-proliferation and apoptosis of the anterior HPAP expressing cells. Thus anterior limb bud HPAP activity is never seen in $Shh^{null,Ssq}$/+ embryos unless a functional anterior ZPA is present, such as the one provided by the *luxate* mutation.

4.3.2 HPAP in homozygous recombinant embryos

Homozygous recombinant embryos generated in section 4.2.3 carried two copies of both the Shh^{null} and *Ssq* alleles, and exhibited HPAP expression in the ZPA of limb buds. Thus the regulatory elements disrupted by the *Ssq* insertion were able to act independently of functional SHH. In other words the *Ssq* regulatory elements act epistatically upstream of *Shh* in patterning the limb field, a result consistent with their suspected role in directly controlling *Shh* expression in the ZPA.

4.3.3 Novel HPAP domains

Two novel domains of HPAP expression were also seen in embryos carrying the *Ssq* transgene at E11.0. The factors responsible for these novel expression patterns of HPAP remain unknown, though it is possible that gene regulatory elements close to the *Ssq* insertion site could be responsible for driving this expression. The best two gene candidates to contain such regulatory elements are *Shh* and *HLXB9*.

Shh is expressed in the ventral neural tube and developing notochord (Echelard et al., 1993b) but this area does not overlap with the expression seen in Fig 4.4 and occurs at E9.5-E10.5. *HLXB9* is a more promising candidate as it has been reported to be expressed in the spinal cord of mice and human embryos (Harrison et al., 1999; Hagan et al., 2000).

However in both organisms the expression occurs earlier than E11.0. Thus the temporal aspect of the novel HPAP spinal expression remains unresolved.

A possible explanation for the anterior limb staining is that it is being driven by the same *Shh* limb regulatory elements highlighted in the *cis-trans* test. It has already been suggested that these elements must be actively repressed in the anterior of the limb and that the *Ssq* insertion acts to disrupt this repression (section 3.6). The embryos in Fig. 4.4 could be exhibiting a small amount of HPAP expression as the repressor begins to shut down the *Shh* limb enhancers in the anterior of the limb. This "leaky repression" model could explain the relative weakness of the expression and its specificity to such a precise time frame.

4.3.4 dHAND and Gli3

The HPAP activity seen in the double recombinant embryos also reveals some insights into the regulation of *Shh* by the transcription factors *dHAND* and *Gli3*. As mentioned in chapter 1 *dHAND* acts antagonistically with *Gli3* in the lateral mesoderm to set up the anterior/posterior axis prior to limb outgrowth (te et al., 2002). Expressed in the posterior area of the prospective limb field, *dHAND* acts to restrict *Gli3* to the anterior of the limb and also stimulates *Shh* expression in the ZPA (Charite et al., 2000). As the limb bud develops *dHAND* continues to be expressed in the posterior domain, and was thought to be involved in a positive feedback loop with *Shh* whereby each maintained the expression of the other as development progressed (Charite et al., 2000). The importance of *Shh* expression to *dHAND* was emphasised when *dHAND's* expression was examined in Shh^{null}/Shh^{null} embryos (te et al., 2002). Instead of a broad expression across the entire posterior of the limb, only a small patch of *dHAND* expression within the posterior of the limb was observed. Additionally due to the reduction in *dHAND* expression, *Gli3* extends its area of expression into the posterior of the limb bud to encompass much of the ZPA (see Fig 4.5).

The double recombinant embryos generated in section 4.5 offered a unique possibility to study the control of *Shh* expressed in the absence of functional SHH, as the *Ssq* HPAP transgene is under the control of the *Shh* limb regulatory elements. As previously noted the HPAP maintained its expression pattern in the ZPA without the presence of functional SHH, and therefore despite *dHAND* having been replaced by *Gli3* in most of the posterior of the limb bud, including the ZPA. Thus it can be inferred from this data that although *dHAND* is essential for the creation of the A/P axis and the ZPA in the limb, it is not required for the maintenance of *Shh* expression in the ZPA. Additionally this data suggests that *Gli3* is unable to downregulate *Shh* expression once the ZPA has developed. Thus although *dHAND* and *Gli3* are undoubtedly important in setting up the initial limb pre-pattern in the lateral mesoderm (te et al., 2002; Charite et al., 2000), it would seem from this

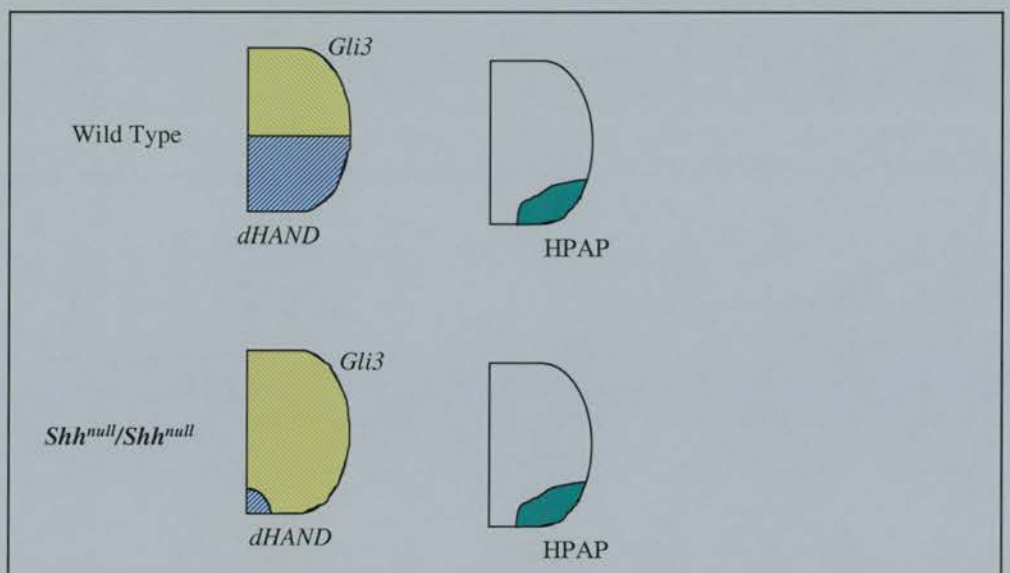


Figure 4.5: A diagrammatic representation of the expression of *dHAND*, *Gli3* and the Ssq HPAP transgene in wild type and Shh^{null}/Shh^{null} embryos. In Shh^{null}/Shh^{null} embryos *Gli3* expression expands to cover the ZPA due to down-regulation of *dHAND*. However the *Ssq* HPAP transgene under the control of *Shh* limb regulatory elements maintains its expression in the ZPA. Thus *dHAND* and *Gli3* are not responsible for maintaining *Shh* expression in the ZPA.

data that once *Shh* is established in the ZPA it is probably unaffected by *dHAND* and *Gli3* expression. Confirmation of the independence of *HPAP* expression from *dHAND* and *Gli3* could come from *in-situs* using *dHAND* and *Gli3* probes on recombinant embryos.

Chapter 5

Comparative sequence analysis

5.1 Introduction

5.1.1 Finding Shh limb regulatory elements

Chapters 3 and 4 concluded that the *Ssq* mutation disrupts the activity of long-range regulatory elements that control expression of *Shh* in the developing limb. However the precise location of these regulatory elements remained unknown. This next chapter records the identification of candidate genomic regions likely to contain *Shh* limb regulatory elements.

Historically the identification of long-range gene regulatory elements such as enhancers or repressors has been a major challenge for researchers. Unlike coding sequences, regulatory elements do not possess well-defined sequence motifs to aid in their discovery. Using cDNA sequences, splice acceptor/donor motifs and genomic features such as CpG islands and RNA polymerase binding sites, computers running gene-prediction programs have made the identification of coding regions relatively trivial. Large-scale sequencing projects such as the human and mouse genome projects have made available vast amounts of genomic sequence, to which gene-prediction software has been applied. Candidate coding regions can then be identified from this "annotated" genomic sequence and confirmed experimentally by any researcher.

Currently, due to their lack of defined sequence motifs, there are no equivalent prediction programs for the identification of gene regulatory elements. Therefore, the identification of candidate regulatory regions has relied on educated guess work, followed by rigorous and time-consuming experimental assays. Genomic sequences upstream, downstream or within introns of coding sequences are traditionally assayed for regulatory activity using reporter constructs. These constructs contain a reporter gene whose protein product can be easily assayed, such as lacZ, coupled to a minimal promoter normally unable to drive expression on its own. Genomic sequences of interest (usually selected on the basis of their proximity to a gene of interest) are then cloned into these constructs, linking

enhancer sequences from these genomic regions to the reporter gene. The modified construct is then used to make transgenic animals or cell lines, and putative enhancer sequences reveal their presence by driving reporter gene expression in an appropriate manner via the minimal promoter (see Fig.5.1). Once a particular genomic region had been demonstrated to exhibit enhancer activity, deletions in the genomic region of the construct help to define the minimal sequence necessary to drive expression. Further refinement of the enhancer sequence then relies on using biochemical assays such as DNAase hypersensitivity studies to define protein-binding sites.

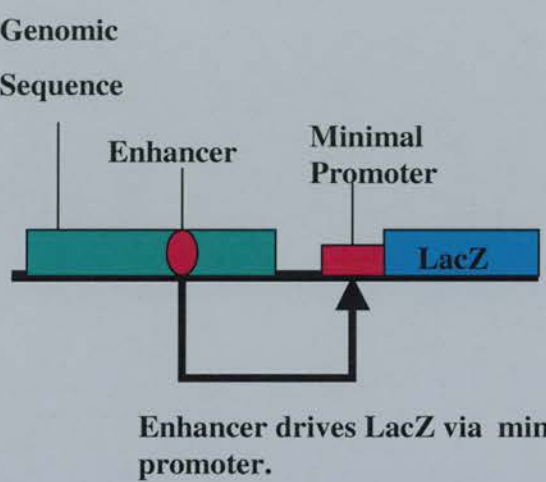


Fig 5.1 Construct to detect Enhancer activity: Enhancer sequences in the genomic region of interest stimulate LacZ expression via a minimal promoter, when incorporated into a transgenic animal or cell line.

This general strategy has been used very successfully to find and define enhancer sequences for many genes (MacKenzie et al., 1997) including those responsible for driving *Shh* expression in the brain and floorplate of developing mouse and zebrafish embryos (Epstein et al., 2000; Muller et al., 1999). However there are several important drawbacks to this strategy that made it unsuitable as an initial step in finding the regulatory elements highlighted by the *Ssq* mutation.

Firstly the reporter assay is incapable of detecting repressor elements, as they will not drive expression of the reporter gene; therefore this strategy was unsuitable for detecting the repressor believed to have been disrupted by the *Ssq* insertion (see Chapter 3). Secondly only relatively small areas of genomic DNA can be assayed at a time, up to 10kb in plasmid constructs or 30-40kb in cosmids. The precise location of the *Ssq* regulatory elements is unknown and they may be scattered over a substantial genomic region, too large to assay with conventional reporter constructs. Although Chapter 3 did highlight several possible enhancer locations based on human and mouse mutations; these were numerous and in the case of the Japanese translocation did not define a definitive genomic area. YAC and BAC constructs could have been used to look at much larger regions but the technical difficulties in handling and assaying such large constructs remained a formidable hurdle. Due to these difficulties it was envisaged that an alternative strategy was needed to find the *Ssq* repressor element and to provide small well-defined candidate regions of genomic DNA that could be assayed for enhancer activity using conventional transgenic reporter constructs. Comparative sequence analysis provided a powerful paradigm that could fulfil these criteria.

5.1.2 Comparative Sequence Analysis of the Ssq locus

Comparative sequence analysis is the process of comparing regions of syntenic genomic DNA between species, and has a long and distinguished history dating back at least to the discovery of bacteriophage promoter and operator sequences (Pribnow, 1975). Comparisons of this sort reveal small regions of conserved non-coding sequences (CNSs) (Hardison, 2000) which, due to their conservation, are believed to perform important functions in the genome such as gene regulation. As species diverge it is presumed that evolution selects against sequence changes in genomic regions essential for survival, thus preserving CNSs between species, unlike non-essential sequences which diverge due to random mutation. Comparative sequence studies of the *HBB* (beta-globin), *Pax6* and *BTK*

(Bruton's tyrosine kinase) (Hardison et al., 1997) loci have revealed multiple CNSs, many of which have been shown to regulate gene expression.

It was hoped that a comparative sequence analysis of the genomic DNA around the *Ssq* genomic region would reveal CNSs as possible candidates for the enhancer and repressor elements discussed in Chapter 3. Such regions could act as a starting point for transgenic reporter studies to find enhancer regions, and would also be good candidates for the *Ssq* repressor. Repressor candidate sequences could then be assayed for mutations in human PPD patients and *Hx* mice.

To carry out any comparative sequence analysis a genomic region of interest must be defined and high-quality DNA sequence from this region obtained from two or more species. During the comparative analysis of the *Ssq* mutation, three regions of interest were defined based on available sequence and the mapped locations of human and mouse limb mutations. The first region of interest spanned genomic DNA from *LMBR1* to *SHH*. The second comprised the entire *LMBR1* gene, from exon one through to seventeen and the third consisted of genomic sequence from *LMBR1* to *KIAA0010*. These three regions contained the mapped locations for all human and mouse limb mutations assigned to 7q36, and together they spanned almost 1.6Mb of human genomic DNA (see Fig 5.2). Thus it was

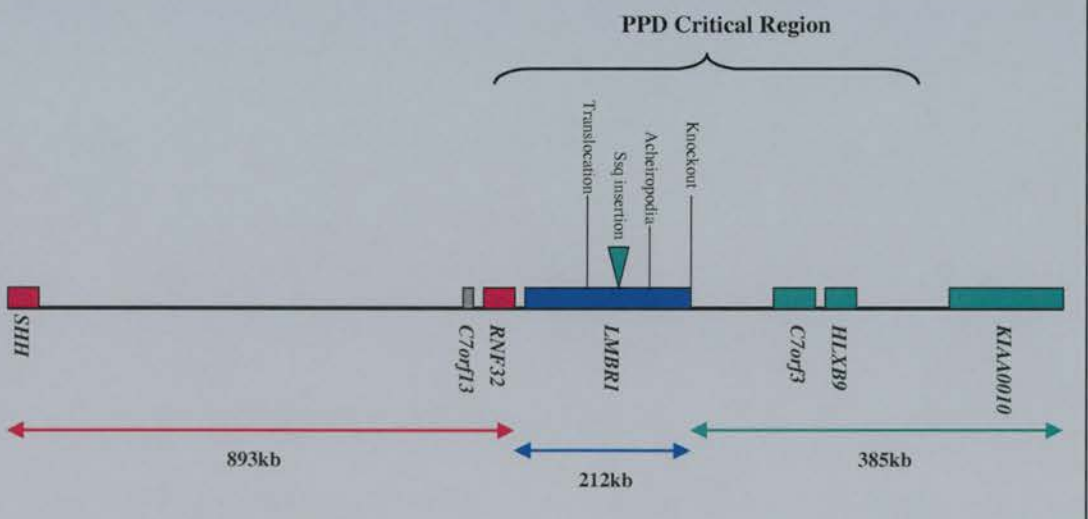


Figure 5.2: Diagram depicting the three genomic regions used during the comparative sequence analysis. Each coloured arrow represents one region.

hoped that all the regulatory elements necessary to drive *Shh* expression in the limb and other tissues were contained within these three genomic regions.

Human to Mouse comparative sequence analysis has been very successful in identifying CNS regions (Hardison et al., 1997), however when examining long stretches of genomic DNA the sheer number identified often precludes the possibility of a systematic transgenic or mutational analysis of them all. For example a study by Loots *et al.*(Loots et al., 2000) compared the sequence of about 1Mb of syntenic mouse and human DNA from chromosomal region 5q31. Comparative sequence analysis revealed 90 CNS regions using the relatively rigorous criteria of at least 75% identity over 100 bp. Extrapolating from this data it, was conceived that in the 1.6 Mb around the *Ssq* locus there could be 150 mouse to human CNS sequences, although some CNSs would be better conserved or in loci disrupted by the various limb mutations, and therefore make better candidate *Shh* limb regulatory elements. It was envisaged that the sheer number of CNS regions between human and mouse would provide too many candidates to be tested.

A way to reduce the number of CNS candidate regions is to carry out further sequence comparisons, with species separated by wider phylogenetic distances (Hardison, 2000); as it is reasonable to assume that non-coding sequences that remained conserved for a longer period of time might play a more fundamental roles in the genome. An excellent species to facilitate this sort of sequence comparison is the teleost fish *Fugu rubripes* (Aparicio et al., 1995). *Fugu* has a compact genome (400Mb): 7.5 times smaller than the human and >90% of it is unique, making genomic sequencing relatively easy (Brenner et al., 1993). Additionally, being a vertebrate *Fugu* has a similar gene repertoire to humans, and there are many parallels between human and fish development, including the expression of *SHH* in the ZPA of limb/fin buds (Akimenko and Ekker, 1995). Finally, of particular importance for comparative sequence analysis is the large phylogenetic distance between *Fugu* and humans: approximately 450myr as opposed to approximately 40myr between mice and humans (Kumar and Hedges, 1998). Therefore in order to get the best possible results

from a comparative sequence approach to identify potential *Ssq* regulatory elements, it was concluded that a 3-way sequence comparison of the relevant genomic regions between Human, Mouse and *Fugu* would be required.

5.2 Results

5.2.1 Obtaining Human and Mouse Genomic Sequence

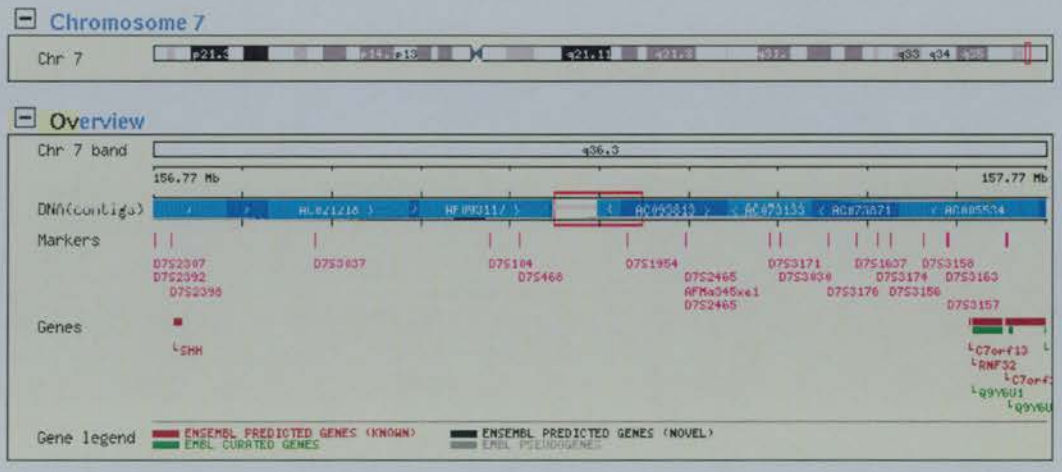
Human and mouse genomic sequence from exon 3 of *SHH* to exon 23 of *KIAA0010* was obtained from the Ensembl Genome browser (http://www.ensembl.org/), human version 7.29a.3 and mouse version 7.3b.3. This server provides a graphical interface for the data generated from the Human and Mouse genome projects. Using the sequence alignment programme BLAST together with *SHH, LMBRl* and *KIAA0010* human cDNAs, the required human and mouse genomic regions were obtained from Ensembl (see Figs 5.3 and 5.4). Large sequence files spanning the genomic region from *SHH* to *KIAA0010* for both mouse and human were downloaded from Ensembl and split into the three regions of interest (see Fig 5.2) using the DNA sequence editing package EMBOSS. These six sequence files (three mouse and three human) were then annotated for gene locations using the sequence alignment programs BLAST and GENEWISE.

5.2.2 Fugu Sequence

Despite the relatively compact genome size of *Fugu* compared to either the mouse or human genomes, there had not until recently been the same emphasis placed into the sequencing of the Fugu genome as had occurred for the mouse/human genome sequencing efforts. Therefore as there was no *Fugu* genomic sequence available around *LMBR1* and the *Ssq* insertion site, it was decided to shotgun sequence a 50kb cosmid (called 16E), that had been isolated from an HGMP Fugu genomic library (a kind gift from Dr. Laura Lettice). The cosmid was known to contain part of the *Fugu Lmbr1* including intron 5 the site the *Ssq* insertion in mouse.

Cosmid 16E DNA was broken up by sonication into fragments between 0.5 and 5kb in size. These fragments were then blunt-ended using T4 polymerase and sub-cloned into the plasmid vector pBluescript that carried an ampicillan resistance gene. These subclones were then electroporated into XL-1 blue *E.coli* cells and spread onto L-Agar plates contaning 10mg/ml ampicillan. Ampicillan-resistant bacterial colonies were then picked into 96-well culture plates and grown overnight. After bacterial growth the culture plates were placed into a robot sample handler and plasmid DNA was generated from each bacterial culture. Using T7 and T3 sequencing primers, the inserts of each plasmid were sequenced and the results collated in the sequence assembly programme CONSED to create contiguous cosmid 16E sequence (see Fig 5.5).

A



B

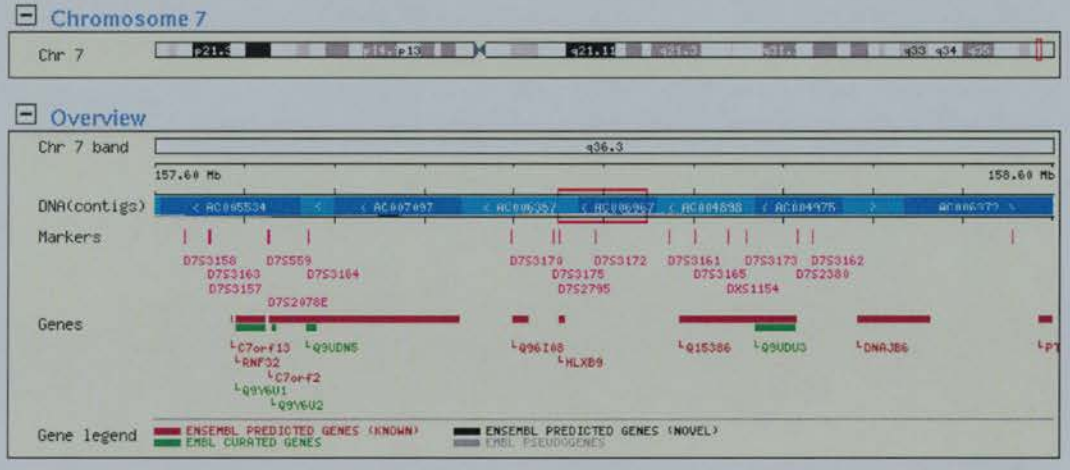
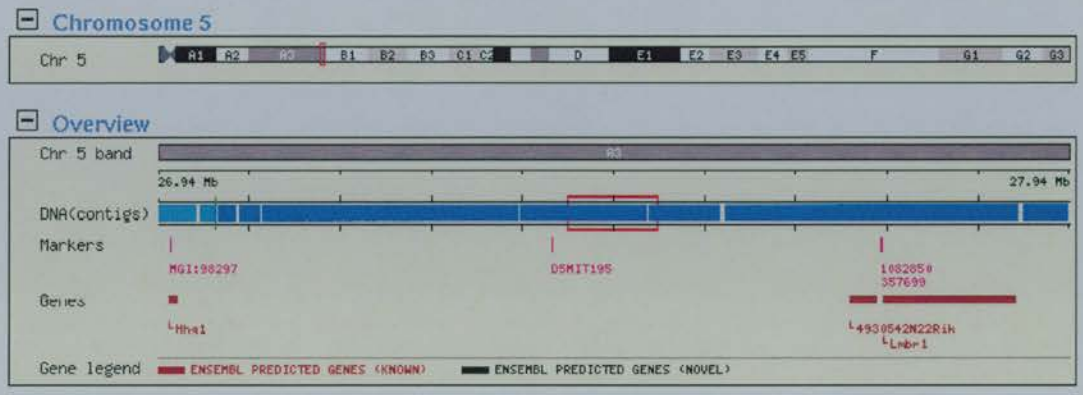


Figure 5.3: The regions of human genomic DNA used in the comparative sequence analysis as viewed in Ensembl. Please note that *C7orf2, Q96I08* and *Q15386* are Ensembl's names for *LMBR1*, *C7orf3* and *KIAA0010* respectively. **A** shows the region from *SHH* to the 3 prime end of *LMBR1* including the gap in sequence between markers D75448 and D751954. **B** shows the human genomic region from *C7orf13* to *KIAA0010*.



B

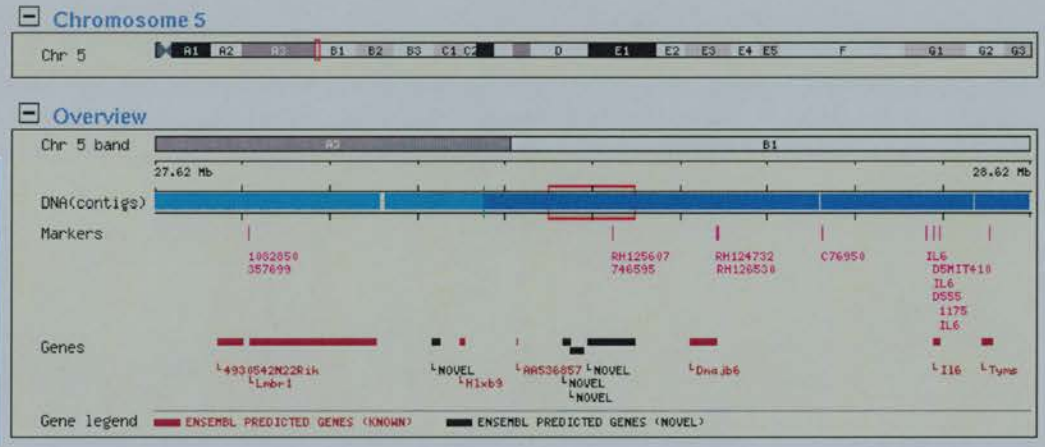


Figure 5.4: The regions of mouse genomic DNA used in the comparative sequence analysis as viewed in Ensembl. Please note that *Hhg1,4930542N22RIK and AA536857* are Ensembl's names for *Shh*, *Rnf32* and *KIAA0010* respectively. **A** shows the region from *SHH* to *LMBR1* including the locations of small gaps in the sequence. **B** shows the human genomic region from *Rnf32* to *KIAA0010*. The novel gene prediction between *Lmbr1* and *Hlxb9* is mouse *C7orf3*. *AA536857* is only a partial mouse EST transcript matched to the mouse genome therefore it appears to be much smaller than *KIAA0010* in the human sequence.

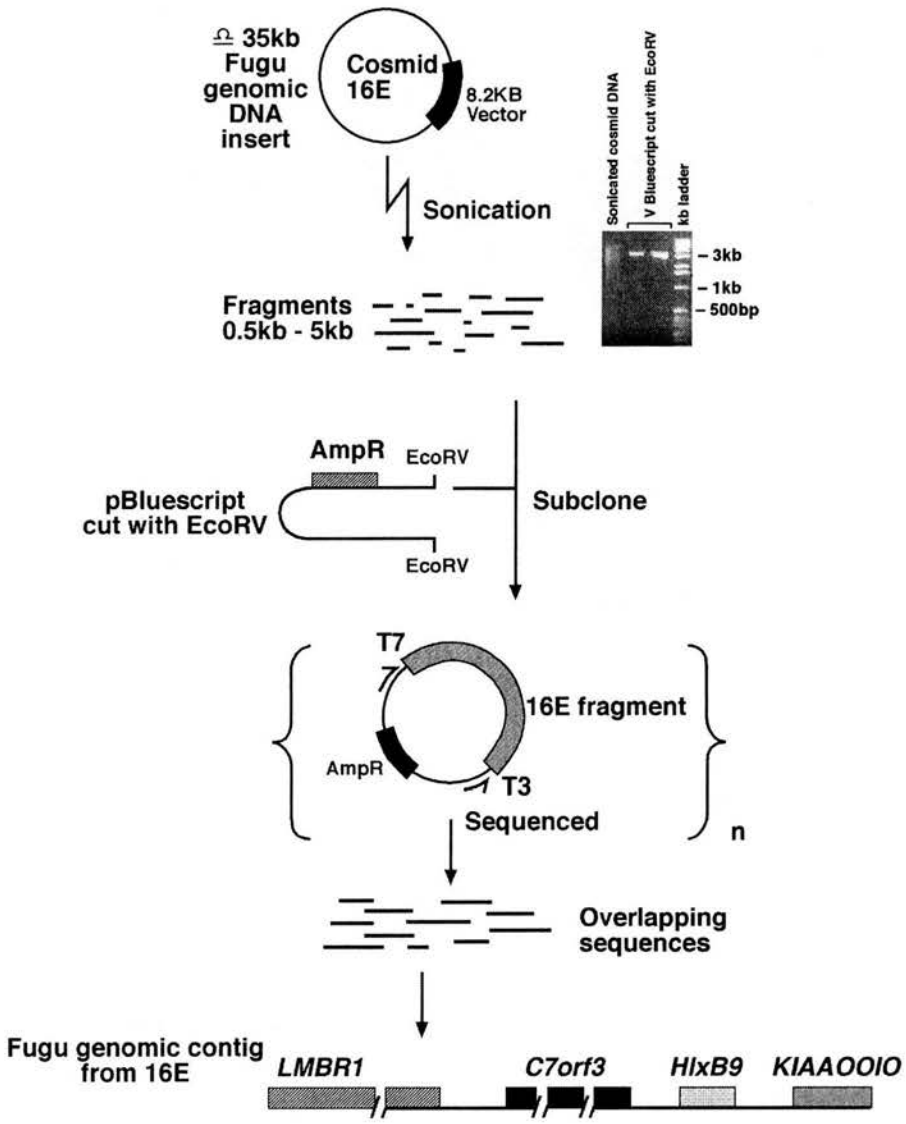


Fig 5.5: Diagram of the strategy used to sequence cosmid 16E. Samples of sonicated 16E DNA and linerised bluescript vector used to generate clones for sequencing are shown in the gel picture above. A map of the contig generated by the sequencing of 16E is also shown.

Three hundred and eight-eight individual sequences were generated ranging in size from 200-700bp, resulting in a total of approx. 160kb of sequence and 3x coverage of cosmid 16E. When assembled, the fragments generated four large contigs that lay within the *Fugu* genomic insert of cosmid 16E and provided sequence information from *LMBR1* exon 11 through to *KIAA0010* exon 10, though three gaps were present: two within *C7orf3* and one in intron 5 of *Lmbr1*(see Fig 5.6). This data strongly suggested that synteny had been preserved between human and *Fugu*, around the *Lmbr1* gene.

Further *Fugu* genomic sequence was made available to us as a result of a collaboration with Dr. Greg Elgar and Debbie Goode (MRC, HGMP, Hinxton, Cambridge). When these sequences were pooled and assembled together with those obtained through our own shot-gun sequencing effort, two large contigs were obtained, which consisted of sequence from *KIAA0010* to just short of *SHH*, though still with a gap in *LMBR1* intron 5. As further data emerged from the public effort to sequence the *Fugu* genome, it was combined with our sequence data, and these gaps were closed resulting in contiguous genomic sequence from *SHH* to *KIAA0010*.

5.2.3 Zebrafish and Tetraodon Sequence

Recently there have been public efforts to sequence the zebrafish and *Tetraodon* (another species of pufferfish) genomes. Although currently at an early stage, large numbers of short genomic sequence reads have been made available and, in the case of zebrafish, a crude genomic assembly has been created. Using BLAST and GENEWISE it was possible to obtain three zebrafish genomic contigs from the HGMP that covered the *Lmbr1* gene, except for gaps in intron 5 and between exons 10 and 16.

At the point of writing this thesis there was no official genome assembly of the *Tetraodon* genome. However, Dr. Martin Taylor (MRC, HGU, Edinburgh) was able to generate a crude *Tetraodon* assembly from *LMBR1* to *SHH* using the equivalent *Fugu* genomic sequence, a novel computer programme called Zeb and the individual genomic

shotgun sequence reads from the public database. Thus, in limited areas of the three genomic regions of interest it was possible to implement a comparative sequence analysis between four or five species.

5.2.4 Sequence Alignments

Finding CNS regions not only requires genomic sequences to compare, but also good computer software to carry out sequence comparisons. Fortunately, two such programmes have been created specifically to carry out this task PIPMaker (Percent Identity Plot Maker) (Schwartz et al., 2000) and VISTA (Visualising global DNA Sequence alignments of arbitrary length) (Mayor et al., 2000).

PipMaker is accessed via a Web server (http://bio.cse.psu.edu/pipmaker/), sequences to be compared are submitted to the web site where they are aligned using a sequence alignment algorithm called BLASTZ. The results of this alignment are then returned to the user, and take the form of a PIP (Percent Identity Plot). PIPs record the position (relative to the base sequence) of regions of identity between the sequences as black lines in the main body of the plot. The higher the recorded position of the line the greater the identity, and the longer the line the greater the size of the conserved region, only regions displaying 50% or more identity are shown. The top axis records pre-determined positions of repeats, CpG islands, and exons in the base sequence. The co-ordinates (lower horizontal axis) represent the nucleotide positions of the base sequence. An example PIP is shown in Fig. 5.6.

Several different parameters can be varied to improve the performance of the BLASTZ alignment programme, particularly when aligning syntenic sequences between species separated by a large phylogenetic distance, such as Human and *Fugu*. These include using a high-sensitivity scoring matrix for BLASTZ, which takes much longer than default settings but is better at finding matches, and the Chaining option, which facilitates the alignment of sequences which differ significantly in size such as equivalent human and *Fugu* genomic regions. PipMaker plots were implemented for all three genomic regions *LMBR1*

exons 1 to 17, *SHH* exon 3 to *LMBR1* exon 1 and *LMBR1* exon 17 to *KIAA0010* exon 1. The human sequence from each region was used as the base sequence; high sensitivity and chaining options were used when comparing *Fugu*, *Tetraodon* and zebrafish sequences. Regions of particular interest from these three plots are highlighted in Figures 5.8, 5.9 and 5.10. Due to their size the full PipMaker plots are contained in the appendix 1.

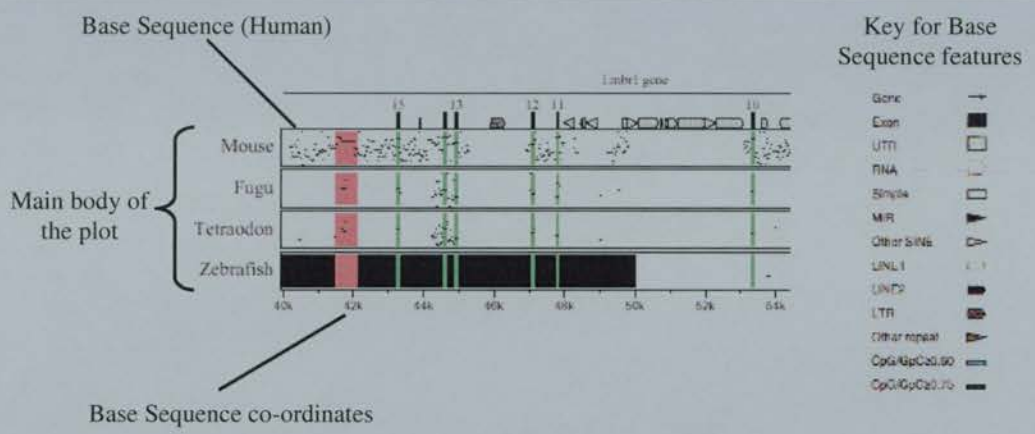


Figure 5.6: An example of a PIP plot. The base sequence (the sequence to which all the other sequences are compared to) is represented by the black line running along the top of the plot. The main body of the plot reveals areas of sequence homology as black lines (pips) for each species examined. The length of the pip represents the length of the homology and the height of the pip represents the extent of the homology. The green stripes represent exons, notice that most exons have a pip on conserved sequence in all the species examined. The red stripe represents a non-coding area of sequence conservation between, human, mouse, zebrafish and *Tetraodon*. The blacked out areas in the zebrafish comparison depict areas where no zebrafish genomic sequence was available.

The VISTA programme uses a sequence alignment algorithm called GLASS, and displays the alignment data as a graph. The x-axis represents the base sequence (human) and the y-axis represents the percent identity, only regions displaying 50% or more identity are shown. Genes and exons are marked above the plot, an example VISTA plot is shown in Fig 5.7. Like PIPMaker, VISTA is accessed via a web server (http://www-gsd.lbl.gov/vista/), where sequences of interest are submitted, and the results returned via e-mail. VISTA was used with all three genomic regions of interest, but was found to be less sensitive than PipMaker in detecting CNSs between human and more phylogentically diverse species such as *Fugu*. This is probably due to the fact that VISTA was not originally designed to compare

sequences over large phylogenetic distances, and therefore does not possess the high-sensitivity and chaining options available in PIPMaker. Due to these shortcomings in VISTA only comparative sequence analysis data produced by PipMaker is displayed in this chapter. An example of a VISTA sequence comparison of the *LMBR1* region is shown in appendix 2. VISTA discovered most of the CNS regions found by PIPMaker, but discrepancies between the two programs are detailed in table 5.1.

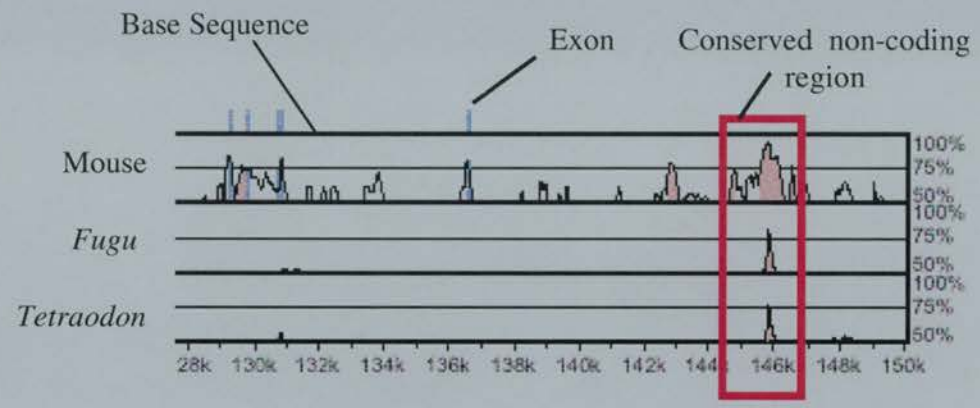


Figure 5.7: An example of a Vista plot. A black line at the top of the line represents the base sequence, and the positions of exons within this sequence are marked as blue boxes. Three graphs below the base sequence depict the level of sequence conservation between the base sequence and mouse, *Fugu* and *Tetraodon* genomic sequences. The y axis of each graph, represents the % identiy, and the x-axis marks the position along the base sequence. A highly conserved non-coding region is marked in red.

5.2.5 Discovery of CNSs

As expected between mouse and human sequences a large number of CNSs were discovered even when using the stringent criteria of 75% homology or more over at least 100bp. Between exon 3 of *Shh* and exon 1 of *KIAA0010* 217 CNSs were found between mouse and human, 16 of which fell within the *Lmbr1* gene. However, when CNSs were defined as non-coding genomic regions conserved between human and mouse (100bp 75% or more) and between human and *Fugu* (100bp 60% or more) the number of CNS regions was reduced to 14. These 14 regions conserved between all three species are hereafter

refered to as 3-sp CNSs (3-species CNSs). Where available the *Tetraodon* and zebrafish genomic showed the same sequence conservation pattern as the *Fugu* genomic sequence.

CNS region:	Location:	Conservation human to mouse:	Conservation human to Fugu:	Detected by Vista?	Est Match?
Cons1	*SHH* int. 1	218bp @ 82%	164bp @ 75%	Yes	No
Cons2	*SHH* to *LMBR1*	359bp @ 87%	102bp @ 68%	No	No
Cons3	*SHH* to *LMBR1*	601bp @ 81%	106bp @ 60%	No	No
Cons4	*SHH* to *LMBR1*	500bp @ 81%	105bp @ 70%	No	No
Cons5	*SHH* to *LMBR1*	457bp @ 82%	106bp @ 77%	Yes	No
Cons6	*SHH* to *LMBR1*	958bp @ 83%	110bp @ 70%	Yes	No
Cons7	*Lmbr1* int. 15	550bp @ 83%	100 bp @ 60%	Yes	No
Cons8	*Lmbr1* int. 9	621 bp @ 87%	120bp @ 80%	Yes	No
Cons9	*Lmbr1* int. 5	1062bp @ 83%	159bp @75%	Yes	No
Cons10	*Lmbr1* to C7orf3	229bp @ 82%	100bp @60%	No	Yes
Cons11	upstream *HLXB9*	1099bp @ 88%	282bp @ 76%	Yes	No
Cons12	*KIAA0010* int. 13	678bp @ 88%	410bp @ 80%	Yes	Yes
Cons13	*KIAA0010* int. 18	633bp @ 82%	109bp @ 70%	Yes	Yes
Cons14	*KIAA0010* int. 18	1053bp @ 88%	238bp @ 76%	Yes	Yes
Exons	*LMBR1*	50-200bp @ 80-92%	50-200bp @ <50-92%	Yes	Yes

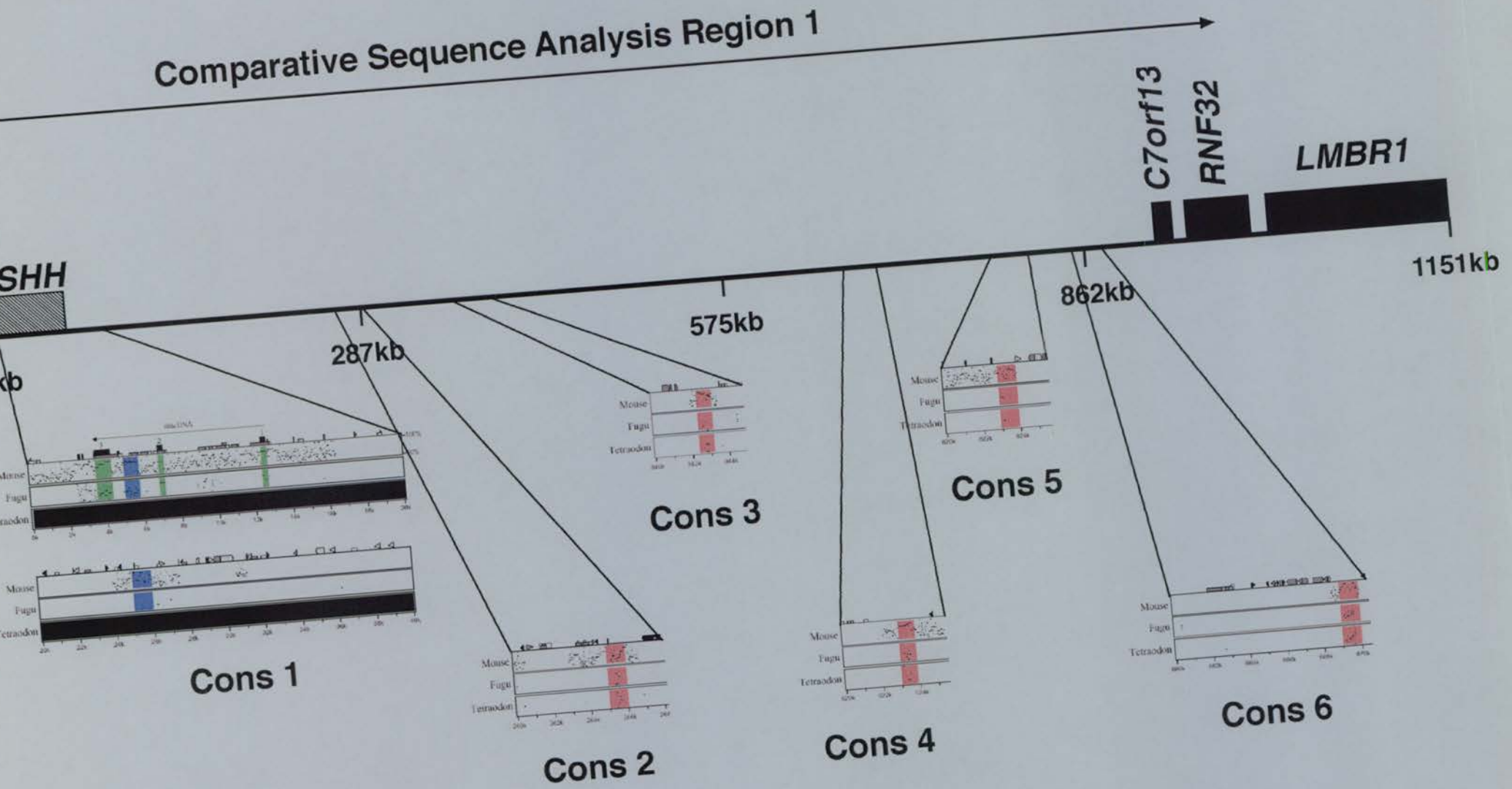
Table 5.1: A summary of the characteristics of the 14 highly-conserved CNSs sequences identified by comparative sequence analysis using PIPMaker. Characteristics of the *Lmbr1* exons are shown in red by way of a comparison.

The 14 3-sp CNSs were named Cons1-14, and their locations and PIPplot profiles are shown in figures 5.8 5.9 and 5.10. A summary of Cons 1-14 features is shown in table 5.1. Cons1-14 sequences were then compared to public sequence databases using BLAST to ascertain if they demonstrated homology to any ESTs or previously characterised regulatory sequences.

Of the 14 3-sp CNS regions four, Cons10, 12, 13 and 14 showed strong homology to mouse or human EST's. These four were discarded from the search for the *SHH* limb regulatory elements as their homology to EST sequences indicated they were probably exons of a transcribed gene, and therefore unlikely to be regulatory elements.

Cons1, a highly conserved region within intron 2 of *SHH,* was revealed by BLAST search to contain two closely linked enhancers of *Shh* which are capable of driving *Shh* expression in the hindbrain/midbrain and floorplate (Epstein et al., 2000). Cons1 was also removed from the search for *Ssq* regulatory elements, as these enhancers did not show any expression within the limb and were therefore unlikely to play a role in the *Ssq* mutation. It

should be noted that a further two previously characterised notochord and floorplate enhancers of *Shh* known to reside approximately 20kb upstream of *SHH* (Epstein et al., 2000) were not defined as highly-conserved CNSs as they did not exhibit significant genomic sequence conservation between human and *Fugu*. The locations of previously characterised enhancers of *Shh* are highlighted in blue in Fig. 5.8.



- PIPMaker results between *SHH* and *LMBR1*. Non-coding genomic regions that are conserved between all the species examined are
and known enhancers are marked in blue.

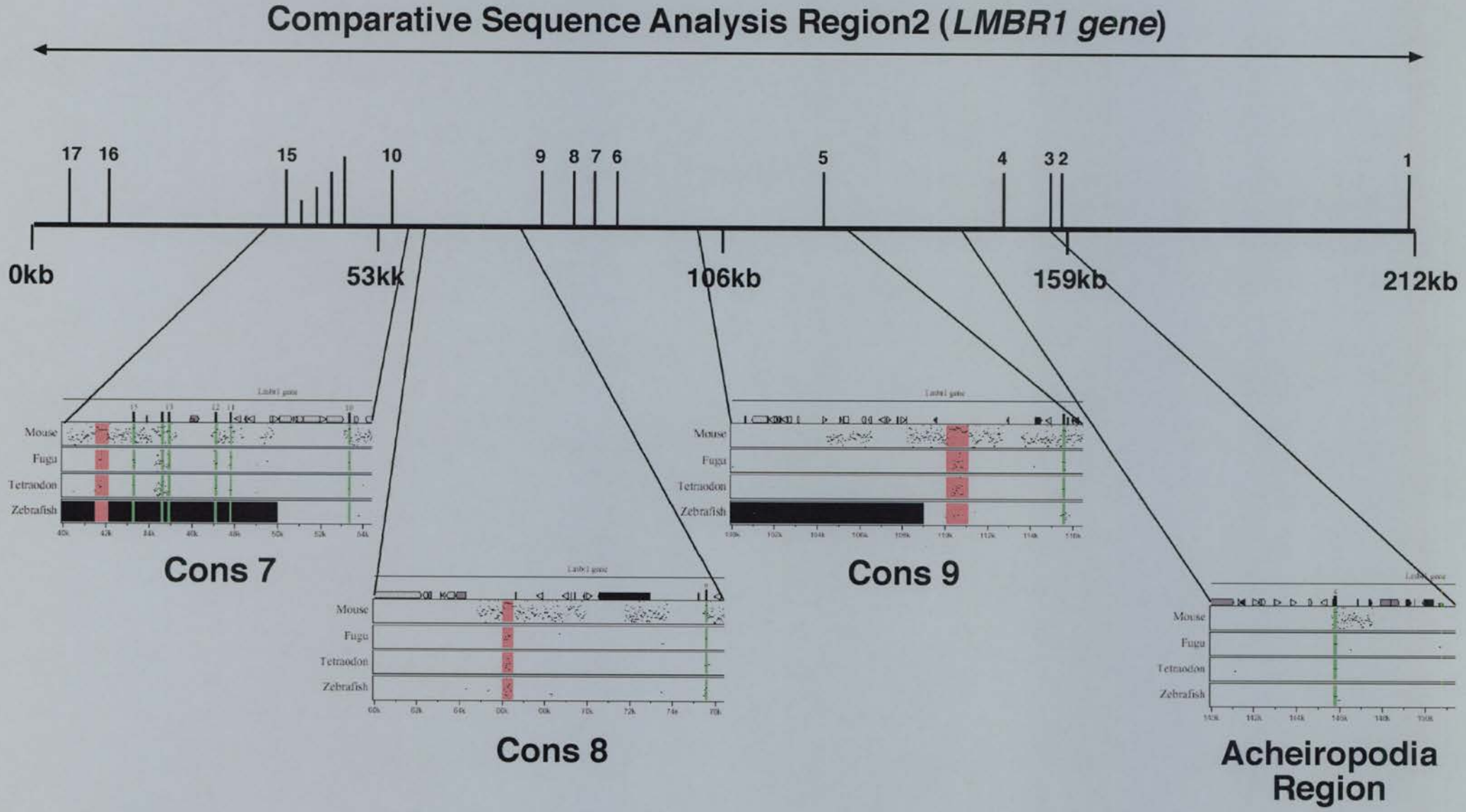


Fig 5.9: Highlights of the PIPMaker results within *LMBR1*. Within the PIP plots non-coding genomic regions that are conserved between all the species examined are highlighted in red, exons in green. The Acheiropodia region is shown, there are no significant areas of non-coding conserved sequence near to *LMBR1* exon 4.

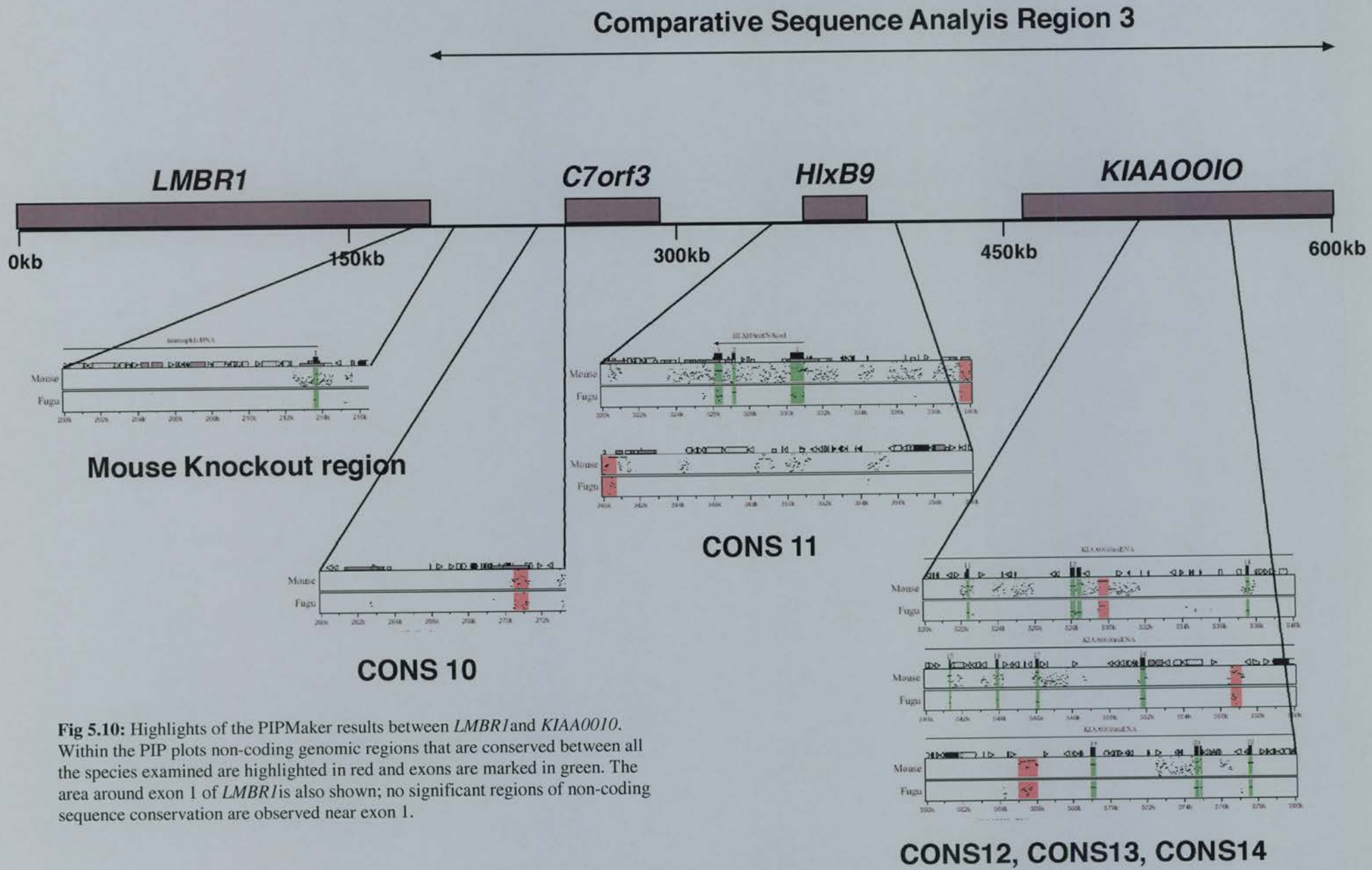


Fig 5.10: Highlights of the PIPMaker results between *LMBR1* and *KIAA0010*. Within the PIP plots non-coding genomic regions that are conserved between all the species examined are highlighted in red and exons are marked in green. The area around exon 1 of *LMBR1* is also shown; no significant regions of non-coding sequence conservation are observed near exon 1.

5.2.6 Discovery of an LMBR1 paralogue, LIMR

To find the human genomic sequence necessary to implement the comparative sequence analysis, a BLAST sequence comparison between the human genome and a *LMBR1* cDNA was done to establish the correct area of the human genome to use. The results from this BLAST established the genomic location of the *LMBR1* gene at 7q36, but also revealed a further strong match to a gene, *lipocalin-1 interacting membrane receptor* (*LIMR*), positioned on human chromosome 12q13.

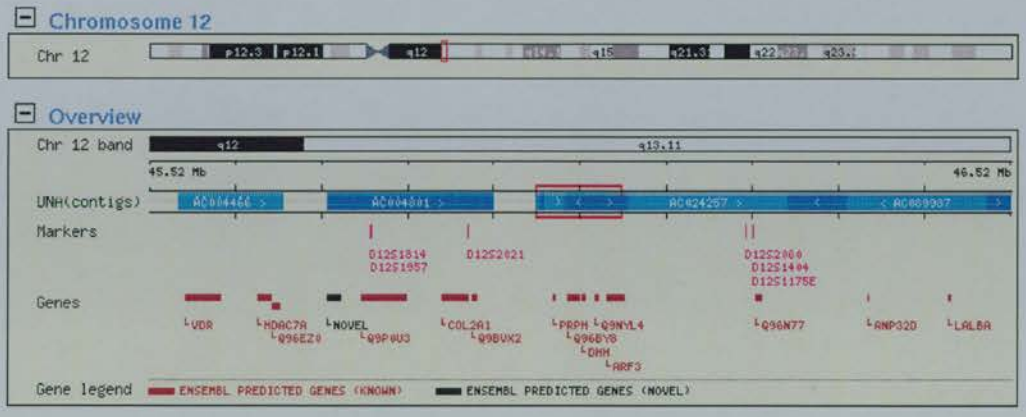
LIMR has 17 exons like *LMBR1,* and is linked to *Desert hedgehog* (*DHH*), a known paralogue of *SHH* (see Fig 5.11). Orthologues of *LIMR* were found in mouse and *Fugu*, and were also linked to *DHH* orthologues in these organisms. Examination of *Drosophilia* and *C. elegans* genomic sequence by Dr. Robert Hill (MRC, HGU, Edinburgh) revealed only one *LMBR1/LIMR*-like gene, which was not linked to *Drosophilia hedgehog (C. elegans* does not have a *hedgehog*-like gene). A CLUSTAL alignment and phylogentic analysis of all the *LMBR* like genes identified is shown in Fig 5.12, and table 5.2 summarises the percent identity and similarity between the *LMBR* like protein sequences.

Gene	% similarity to Human Lmbr1	% identity to Human Lmbr1	% gaps compared to Human Lmbr1
Mouse *Lmbr1*	98	95	0
Fugu Lmbr1	88	79	0
Human *LIMR*	76	58	2
Mouse *LIMR*	75	58	2
Fugu LIMR	71	54	6
Drosophlia Lmbr1	58	38	11
C.elegans Lmbr1	40	22	35

Table 5.2: The results of an amino acid sequence comparison of the *Lmbr1* and *LIMR* genes to human *Lmbr1*. The % gaps refers to differences in size between the human *Lmbr1* and the genes examined.

This data reveals that *LMBR1* has orthologs in *Drosophlia* and *C.elegans* as well as mouse and *Fugu*. However, in vertebrate species there are two *LMBR*-like genes, which due to their amino acid similarity are probably paralogues.

A.



B.

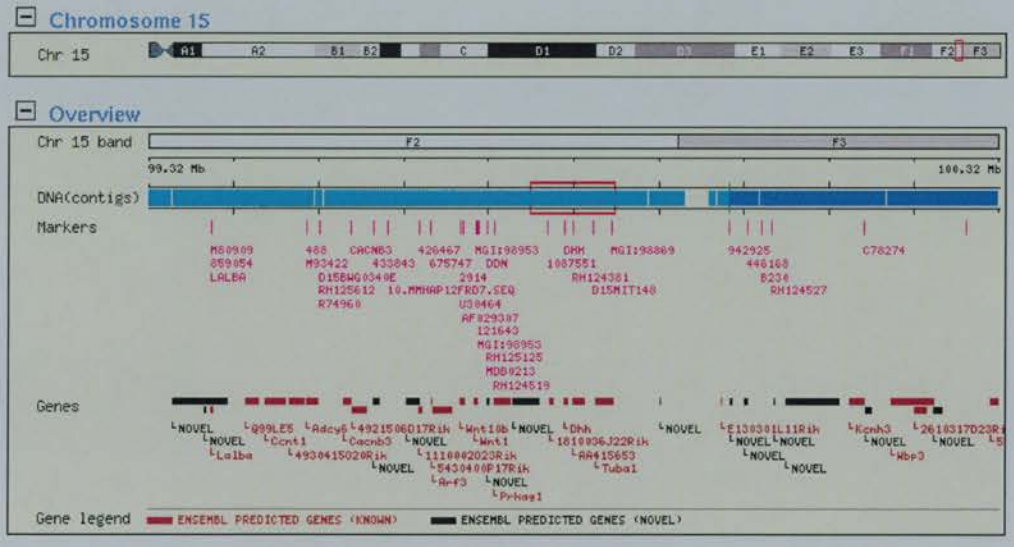
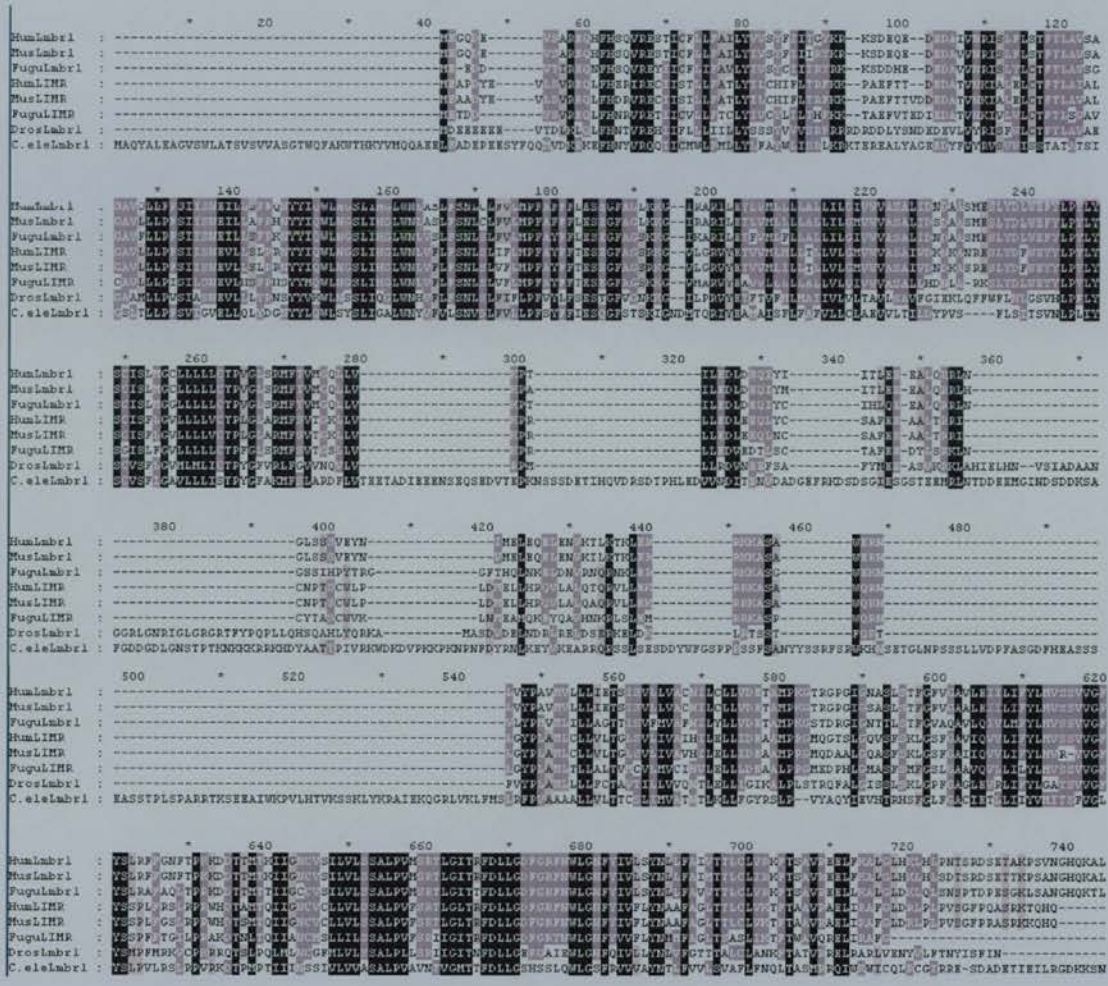


Figure 5.11: **A** The region of human genomic DNA that contains *LIMR* and *DHH* as viewed in Ensembl. Please note that *Q96BY8* is Ensembl's name for *LIMR*. **B** The mouse genomic that contains *Limr* and *Dhh* as viewed in Ensembl. The gene *AA415653* is Ensembl's name for *Limr*.

A.



B.

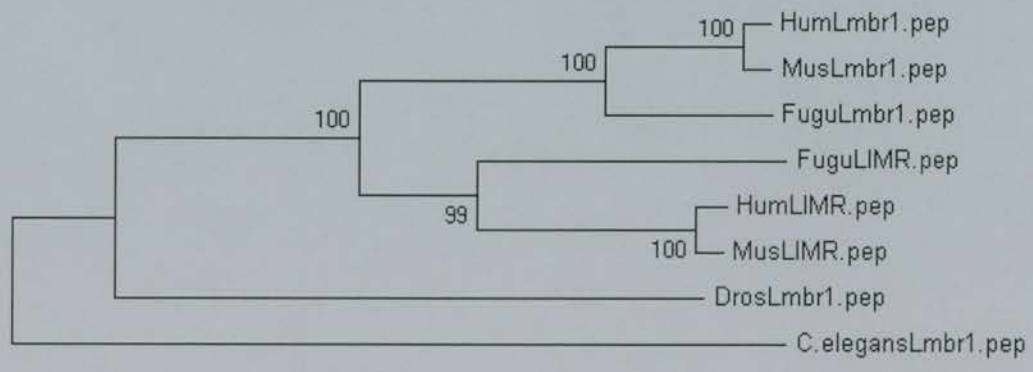


Figure 5.12: A Records the results of a clustalW line-up of human, mouse, *Fugu*, *Drosophila* and *C.elegans Lmbr1/LIMR* protein sequences. **B** A phlylogenetic tree of the family of *Lmbr1* genes, that includes all the protein sequences used in the clustalW line-up. The phylogenetic analysis was conducted with 1000 bootstrap replications of the neighbour-joining method. From this analysis it is evident that the *Lmbr1* genes from mouse, human and *Fugu* are more closely related to each other than their corresponding *LIMR* gene.

Although the coding regions and structure of *LIMR* and *LMBR1* are very similar in all species examined, their genomic contexts differ dramatically. In human, mouse and *Fugu* *LMBR1* has much larger introns and is spread over a greater genomic area than *LIMR*, and the genomic distance between *LMBR1* and *Shh* is much larger than between *LIMR* and *Dhh*. Although an increase in repetitive DNA accounts for some of these differences, much of the additional genomic DNA is unique (see table 5.3).

Species	Genomic Region	Length	% repeat DNA	% unique DNA	Unique DNA
Human	*Lmbr1*	212kb	60	40	85kb
Mouse	*Lmbr1*	145kb	40	60	87kb
Fugu	*Lmbr1*	22kb	1	99	22kb
Human	*LIMR*	13kb	27	73	9kb
Mouse	*LIMR*	13kb	21	79	10kb
Fugu	*LIMR*	6kb	0	100	6kb
Human	*Lmbr1 to Shh*	1151kb	38	62	713kb
Mouse	*Lmbr1 to Shh*	945kb	39	61	576kb
Fugu	*Lmbr1 to Shh*	80kb	1	99	79kb
Human	*LIMR to Dhh*	24kb	35	65	16kb
Mouse	*LIMR to Dhh*	30kb	28	72	22kb
Fugu	*LIMR to Dhh*	15kb	1	99	15kb

Table 5.3: This table records the genomic size (rounded up to the nearest kb) and % repeat DNA (to the nearest whole number) of the *LMBR1* and *LIMR* genomic regions. Characteristics of the intervening genomic DNA between *LMBR1*/*LIMR* and their respective hedgehog genes is also shown.

5.2.7 LIMR Comparative Sequence Analysis

Due to *LIMR's* similarity to *LMBR1* and linkage to *Dhh,* it was thought that LIMR might also contain regulatory regions, perhaps acting on *Dhh*. To ascertain if this were true, a comparative sequence analysis was implemented using syntenic human, mouse and *Fugu* sequences that contained the *LIMR and Dhh* coding regions plus all the intervening genomic sequence. The results of a three-species PIPMaker analysis of this region are shown in Fig. 5.13. 11 CNS regions (minimum 75% homology over 100bp) were found between human and mouse but no CNS regions were found between all three species.

Figure 5.13: PIP-plot comparison between human, mouse and *Fugu LIMR* to *DHH* genomic sequences. The base sequence is human, the numbers at the bottom of the plot correspond to distance along the human sequence in 1000's of base pairs. Exons of *LIMR* and *DHH* are marked in green, human repeat elements are marked along the top of the plot. Regions of homology between mouse or *Fugu* sequence are depicted as black horizontal lines in the body of the plot. The length of the line depicts the size of the region of homology, the higher up the plot the more homologous the region. As can be seen from this plot the exon sequences of the two genes are well conserved, but no significant regions of non-coding homology are conserved across all three species.

5.3 Discussion

5.3.1 Limb mutations and CNS regions

The purpose of this comparative sequence analysis study was to identify candidate genomic regions that could be long-range regulators of *SHH* expression during limb development. It was concluded in chapters 3 and 4 that the disruption of these element's activity was the cause of a variety of human and mouse limb mutations (including *Ssq*) that mapped to human 7q36, and its syntenic region on mouse chromosome 5. The criteria used to identify a potential regulatory element were that it, firstly, be non-coding, and, secondly, be highly conserved across a large phylogenetic distance. However, to be involved in the regulation of *SHH* during limb development it is reasonable to assume that the candidate regulatory sequence must also be located in a genomic region associated with the limb mutations mapped to human 7q36 and mouse chromosome 5.

The comparative sequence analysis of human, mouse and *Fugu* genomes between *SHH* and *KIAA0010* proved to be successful in identifying nine CNSs as candidate regulatory sequences. But only four of these CNSs were contained within genomic regions associated with limb defects and the disruption of *SHH* limb expression (see Fig 5.14).

The most promising candidate of the four, Cons9 was the only highly conserved CNS located in a genomic locus disrupted by the *Ssq* insertion event. As described in chapter 1, the *Ssq* insertion results in a partial duplication of *Lmbr1* intron 5 (Lettice et al., 2002), a genomic region that includes Cons9 (see Fig). With regards to the highly conserved CNS regions, the only difference between *Ssq* and wild type mice, is that *Ssq* mice contain extra copies of Cons9. Thus, it is highly likely that Cons9 is the *Shh* limb regulatory element believed to be disrupted by the *Ssq* insertion event, though it is unclear at this stage how a duplication of Cons9 could result in ectopic anterior *Shh* expression.

Cons7 and Cons8 are located in *LMBR1* introns 15 and 9 respectively; both loci are within the critical genomic regions associated with a variety of human and mouse limb defects. Thus both could be regulatory regions involved in the long-range regulation of *SHH* in the developing limb.

Cons11 is located approximately 10kb upstream of *HLXB9*, just within the human

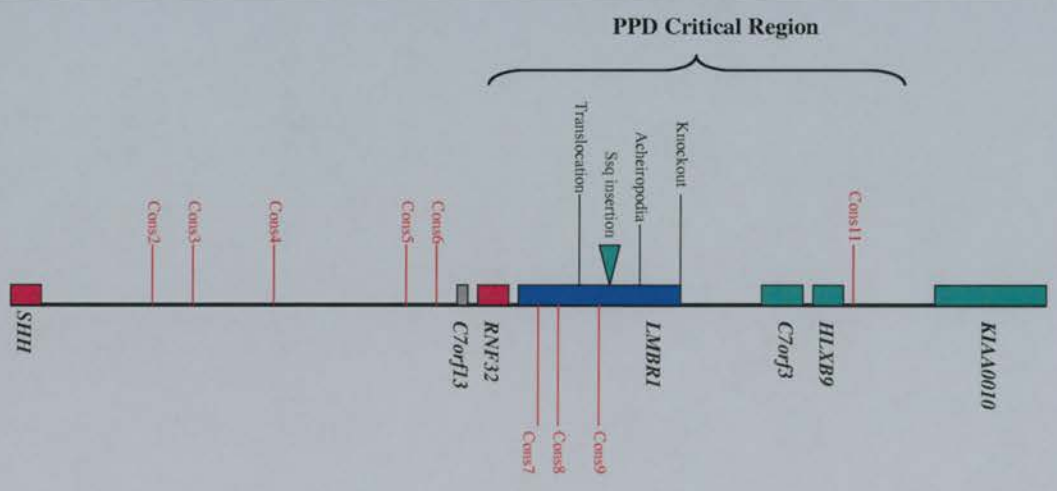


Figure 5.14: A cartoon depicting the positions of Cons 2-9 and Cons 11 in relation to human and mouse limb mutations. Cons 7-9 and Cons 11 lie within the PPD critical region, but only Cons 9 is disrupted by the *Ssq* mutation.

PPD critical region but outside the genomic region associated with the mouse *Hx* limb mutation. Although it is possible that Cons11 could be a long-range limb regulator of *SHH*, its proximity to *HLXB9* suggests that it is more likely to be a regulator of *HLXB9* expression.

Further experimental assays are required to establish if any of the candidate regulatory sequences are indeed long-range limb regulators of *SHH*. The results of transgenic assays to establish if Cons8 and 9 have enhancer activity are described in Chapter 6, and information regarding an analysis of Cons9 for mutations in PPD patients is discussed in Chapter 7.

5.3.2 CNS regions between SHH and LMBR1

The remaining five highly-conserved genomic regions Cons2 to 6 were all located between *Shh* and *C7orf13*. Due to their high sequence conservation between species it is possible that they act as regulatory elements, but as they are located outside the genomic region associated with limb mutations they are unlikely to involved in limb development. Clues to a possible role for Cons2-6 come from a further set of human mutations that cause holoprosencephaly, a common developmental defect of the forebrain and mid-face (Roessler and Muenke, 2001).

Haploinsufficiency of *SHH* has been proposed as a cause of holoprosencephaly (HPE) in humans (Nanni et al., 1999), due to *SHH's* central role in the patterning of the Central Nervous System (Echelard et al., 1993a) and the fact that $Shh^{-/-}$ mice exhibit cyclopia (Chiang et al., 1996), a severe form of HPE. Additionally heterozygous nonsense and missense mutations have been found in *SHH* of HPE patients (Roessler et al., 1996). However a further subset of HPE patients have been described that exhibit translocations 15-250kb downstream of *SHH* (towards C7orf13), but display no *SHH* coding mutations (Belloni et al., 1996), suggesting that a "position effect" has an important role in the aetiology of HPE. Such a position effect could be explained by the separation of *SHH* forebrain enhancers from the *SHH* coding region by the translocation events. This scenario

would result in a down-regulation of SHH during development of the forebrain, causing HPE. Cons2-6 are excellent *SHH* fore-brain enhancer candidates since they are highly conserved non-coding regions and are separated from the *SHH* coding region by the HPE translocations (see Fig 5.15).

Obviously transgenic assays would need to be carried out to confirm that Cons2-6 have enhancer activity consistent with expression in the fore-brain. It is also conceivable that

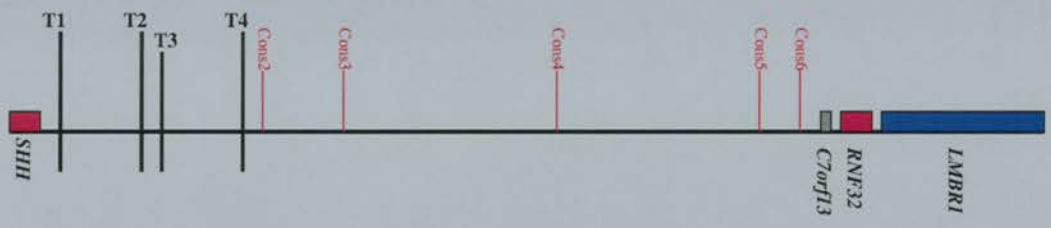


Figure 5.15: Cartoon showing the relative positions of Cons 2-6 and 4 translocations (T1-4) believed to cause HPE. All of the translocations result in the separation of Cons2-6 from *SHH*. Thus if Cons 2-6 contain *SHH* fore-brain regulatory elements there separation could result in a down-regulation of SHH in the forebrain and an HPE phenotype.

Cons2-6 might be responsible for regulating other areas of *SHH* expression, such as in the gut or hair follicles.

5.3.3 Acheiropodia, and Lmbr1 knockout mice

At the end of Chapter 3 it was proposed that the Acheiropodia mutation, and the *Lmbr1* knockout removed limb enhancers of *Shh*, however this comparative sequence study has not revealed any highly-conserved CNS regions in the genomic loci removed by these mutations (see Figs 5.7, and 5.8). Therefore, it is difficult to reconcile these mutations with the suggestion that all limb mutations in the *LMBR1* genomic region act by disrupting limb regulatory elements of *SHH*. However, it should be borne in mind that comparative sequence analysis is not infallible and sequence conservation between humans, mice and fish does not necessarily predict the locations of all the regulatory regions present in any particular genomic region. Indeed within this study a previously characterised *SHH*

floorplate/notochord enhancer was not predicted by comparative sequence analysis due to the enhancer's lack of conservation in *Fugu*. However, it must be stated that, at this stage, it is not possible to reconcile the *Lmbr1* knockout and Acheiropodia mutations with the *Ssq* cis-acting regulatory model proposed as a result of the *cis-trans* test in Chapter 3, although an effort with additional data will be made in Chapter 7.

5.3.4 LMBR1 and LIMR

Data in Section 5.5 demonstrated a strong similarity between the protein sequences of *LMBR1* and *LIMR*. *LIMR* and *LMBR1* orthologues were found in human, mouse and *Fugu*, but only a single *LMBR1/LIMR* orthologue was found in *Drosophlia* and *C.elegans*. Phylogenetic analysis of available *LMBR1* and *LIMR* sequences revealed that the *LMBR1* orthologues showed greater similarity to each other than to their corresponding *LIMR* gene. The simplest explanation for this is that *LMBR1* and *LIMR* are paralogues, i.e. derived from a gene duplication event. This event probably occurred after the divergence of the Nematodes/Arthropods and Chordates as only a single *LMBR* like gene is present in *C.elegans* and *Drosophlia*, but before the divergence of the Actinopterygii fish (the ray-finned fishes including *Fugu* and zebrafish) from the Sarcopterygii lineage (the lobe-finned fishes including the coelacanth and lungfish, from which birds and mammals are derived), as *LMBR1* and *LIMR* are present in human, mouse and *Fugu* genomes. The linkage of *LMBR1* and *LIMR* to *Dhh* and *Shh* (paralogous genes), respectively, in human, mouse and *Fugu* suggests that the duplication was not confined to a single ancestral *LMBR* gene, but incorporated a larger genomic region.

Thirty years ago it was proposed that the vertebrate genome had undergone several rounds of genome duplication (polyploidisations) (Ohno S., 1970). Evidence in support of this has mainly come from the observation of large numbers of duplicate vertebrate genomic regions (paralogons) (Abi-Rached et al., 2002; McLysaght et al., 2002)such as the *Hox* gene cluster, the major histocompatibility complex and *LMBR/Hedgehog* region. Clearly,

paralogons could have arisen as a result of duplications of portions of the genome. But the sheer number of paralogons tends to argue in favour of a few whole genome duplication events rather than a multitude of smaller duplication events (McLysaght et al., 2002). Currently, it is believed that two genome duplications occurred during the early evolution of vertebrates. The first probably pre-dates the Cambrian explosion, before Chondricthyes (sharks, dogfish etc.) split from the Actinopterygii/Sarcopterygii lineages, and the second possibly dates back to the early Devonian period (Meyer and Schartl, 1999). Recent data suggests that the Actinopterugii lineage then underwent a third round of duplication to produce eight copies of the original deuterostome genome (Amores et al., 1998). Genome duplication is believed to be an important mechanism for functional innovation during evolution (Graham, 2000). Although many duplicated genes would rapidly become redundant after a duplication event, eventually being lost from the genome, many others may have evolved novel, though related, functions.

The classic example of gene duplication followed by selective gene loss and gain of novel gene function is a collection of linked genes called the *Hox* cluster, believed to have originated from an ancient cluster of 13 linked genes (Holland and Garcia-Fernandez, 1996). *Hox* genes are implicated in the control of embryonic axial patterning of probably all animal species. Invertebrate chordates such as Amphioxus have one cluster, mammals have four and zebrafish have seven (Holland and Garcia-Fernandez, 1996) (Amores et al., 1998). The duplicated *Hox* clusters in mammals show substantial gene loss, to the extent that no one cluster contains a full complement of 13 paralogous genes (Holland and Garcia-Fernandez, 1996). However the addition of extra *Hox* genes has been suggested to facilitate the evolution of more complex vertebrate body plans, and *Hox* genes are certainly expressed in more derived structures such as the brain and limb (Holland and Garcia-Fernandez, 1996).

These genome duplication events are the most likely mechanism by which an ancestral *LMBR* gene could have been duplicated to form *LMBR1* and *LIMR*. The most likely scenario is that after the first genome duplication, one copy of the duplicated *LMBR* was lost

from the genome. The second duplication event produced *LMBR1* and *LIMR* which quickly gained different functions, maintaining their presence in mouse, human and *Fugu* genomes.

LIMR has been thought to play a role in the innate immune response as it has been proposed as a receptor for lipocalin-1 (Lcn-1) (Wojnar et al., 2001), whose roles include stabilising the lipid film of human tear fluid, and possibly the scavenging of potentially harmful lipophylic compounds. Currently the function of *LMBR1* is unknown, though suggestions have been made that it may play a role in limb development (Clark et al., 2000; Clark et al., 2001; Ianakiev et al., 2001)(see Chapter1). The two proposed functions for these paralogous genes seem inconsistent, one a developmental gene and the other involved in immunity. Although paralogous genes are envisaged to gain independent functions in order to survive selection, the functions of the two paralogues tend to be related, i.e. both tend to be involved in similar process such as anterior/posterior patterning (*Hox* cluster) or the immune system (human MHC paralogues). In the light of the *cis-trans* test (Chapter 3), which suggested that *Lmbr1* is incidental to the mutations that cause PPD, it seems more likely that *LMBR1* might play a role in innate immunity similar to *LIMR* rather than acting to direct limb development.

5.3.5 Differences in Genomic Context

LMBR1 and *LIMR* display a high degree of similarity in their coding regions and intron/exon structure in all the vertebrate species examined so far. However other aspects of these paralogues genomic context display some striking differences. The most obvious is the difference in intron size between paralogues, for example the human *LMBR1* coding region covers a genomic area almost 20 times the size of the human *LIMR* coding region. Additionally, although both paralogues are linked to *hedgehog* genes, the distances between them are not equal as human displays an almost 50 fold increase in DNA between *LMBR1* and *SHH* relative to *LIMR* and *DHH*. This pattern of increased genomic DNA at the *LMBR1* loci is repeated in all three species examined.

Although an increase in repetitive DNA at the *LMBR1* loci accounts for some of these differences, much of this additional DNA is unique. Currently it is impossible to ascertain whether the difference in loci size is due to the *LMBR1* loci gaining DNA or the *LIMR* loci losing DNA. Nor is it possible to give a definitive answer as to why this difference in loci size should occur between such closely related paralogues. However one possible speculation concerns the *Shh* regulatory elements highlighted by the *cis-trans* test in Chapter 3 which are believed to lie within the *LMBR1* locus.

Unlike *LMBR1, LIMR* does not contain any 3-sp CNSs between mammels and *Fugu* and is therefore unlikely to contain vital gene regulatory elements. It is conceivable that the evolution of *Shh* regulatory elements within *LMBR1* has facilitated the increase in size of the *LMBR1* locus relative to *LIMR*. The *LMBR1* locus probably gained *Shh* limb regulatory elements in the Actionpterygii/Sarcopterygii lineage the split with Chondricthyes (sharks, dogfish, etc.), as sharks are not believed to exhibit *Shh* expression in the limb (Tanaka et al., 2000). Additionally, the plethora of highly-conserved non-coding sequences within *LMBR1* and towards *SHH* suggests that this region gained these sequences after the second vertebrate genome duplication, as no similar sequences are present within the *LIMR* region.

It is possible to envisage that a series of transposon or recombination events may have deposited regulatory regions from other areas of the genome into the *LMBR1* locus, whereupon they evolved to regulate *Shh* limb expression. As the regulatory elements gained essential functions, evolution would have selected against recombination events resulting in their loss. Therefore it would be less likely for the removal of DNA to occur within the *LMBR1* locus. However recombination events that added DNA probably would not interfere with the activity of the regulatory elements, as regulatory elements tend to be distance independent, thus the *LMBR1* locus would be more likely to grow than shrink. Without regulatory elements to positively select for recombination events that add DNA, the *LIMR* locus was unlikely to grow to the same extent as *LMBR1*.

Further insights into the evolution of the *LMBR1* and *LIMR* loci would be provided by examination of equivalent loci in other species, particularly the cartilaginous fishes (sharks, dogfish etc.). Currently little is know about shark genomic DNA, or how many genomic duplications it has undergone (though probably at least one). Though from an evolutionary perspective it would be of value to know how many *LMBR* paralogs shark possesses, and secondly if any are linked to *hedgehog* genes. However without *Shh* expression in the limb it is reasonable to predict that shark genomic DNA does not contain the *LMBR1* regulatory elements. Therefore, a sensible prediction would be that any shark *LMBR*-like loci would be condensed, looking more like *LIMR* and may not be linked to a *hedgehog* gene.

Despite a lack of *Shh* in the limb, and therefore presumably a lack of *LMBR1* regulatory elements, the cartilaginous fish still develop pectoral and pelvic fins. The obvious question therefore is why did regulatory elements evolve to drive *Shh* expression in the developing limb bud. One possible suggestion that has been made is that *Shh* is involved in "freeing the fins" and establishment of a proximal-distal limb axis (Tanaka et al., 2002). In *S.canicula* (dogfish), as in many other cartilaginous fishes, the metapterygium (main long-bone of the fin) lies proximal and parallel to the body wall (Tanaka et al., 2002). However in higher vertebrates the long-bones grow out perpendicular from the body wall (see Fig 5.16). The advent of the *LMBR1* regulatory elements driving *Shh* limb expression may have facilitated this re-orientation of the limb long bones, thereby ushering in a vital step in vertebrate evolution by creating the potential for the formation and patterning of more complex proximal-distal limb structures.

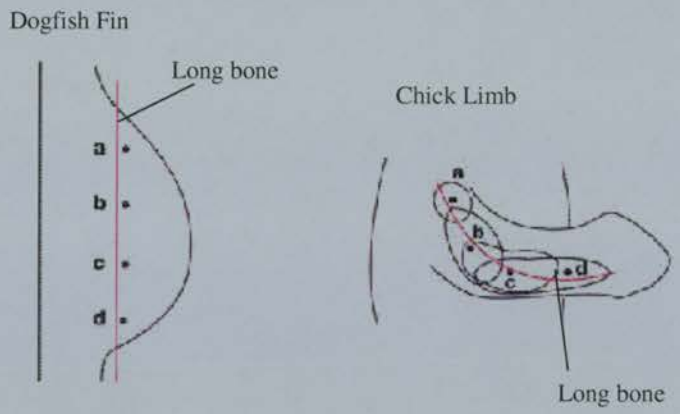


Figure 5.16: A cartoon of a dogfish fin and chick limb taken from Mikiko Tanaka *et al.* The long bones are represented as red lines. Dogfish long bones develop parallel to the body wall but chick long bones grow out perpendicular to the body wall. The outgrowth of the limb is perceived as a major evolutionary step in the development of tetrapod limbs, and is thought to be driven by *SHH* expression in the limb.

Chapter 6

Transgenic analysis

6.1 Introduction

Chapter 5 identified several regions of non-coding genomic DNA that were highly conserved between human, mouse and *Fugu* (3-species CNSs). Three of these CNSs were contained within the critical regions for both human PPD and the mouse mutation *Hx,* and therefore were excellent candidates as potential regulatory elements of *Shh* limb expression. We selected two of these CNSs, Cons8 and Cons9, for further study, based on the fact that they were the most highly conserved and closest to the *Ssq* insertion site.

The *HPAP* expression pattern in *Ssq* limbs suggested the presence of regulatory elements that were capable of driving gene expression that mirrored that of *Shh* in the limb. In the light of the *cis-trans* test in chapter 3 it seemed likely that these elements were disrupted by the *Ssq* mutation, and were responsible for the *Ssq* phenotype. As an initial step in determining if Cons8 and Cons9 were involved in regulating *Shh* expression in the limb, it was deemed prudent to establish if they were capable of driving gene expression in a similar pattern to *HPAP* in *Ssq* limb buds. If this were the case then it would seem reasonable to suppose that the CNS driving expression was in fact a limb enhancer of *Shh*.

With this aim in mind, a strategy was devised whereby Cons8 and Cons9 would be cloned into a, *LacZ* minimal promoter reporter construct, and then assayed for enhancer activity using transgenic embryos harvested at appropriate stages of development. As has already been mentioned in chapter 5 this traditional method of assaying genomic sequence for enhancer activity has been very successful in identifying a number of enhancer regions. Dr. Laura Lettice and Mrs Lorna Marshall (MRC Human Genetics Unit) carried out the cloning and transgenic assaying of Cons9; the equivalent work on Cons8 was carried out by myself and is described in the rest of this chapter.

6.2 Results

6.2.1 The Cloning of Cons8

Cons8 consists of 621bp of conserved genomic DNA (human to mouse); in the centre of which is120bp that is conserved between human, mouse and *Fugu*. To ensure the best possible chance of obtaining a functional enhancer sequence from Cons8 it was deemed prudent to clone all 621bp conserved between human and mouse plus 450bp upstream and downstream of Cons8; a total genomic fragment of 1521bp. PCR primers were designed against mouse genomic sequence upstream and downstream of Cons8, and contained non-homologous ends that incorporated Hind III sites; to facilitate subsequent cloning of the PCR product (see Fig 6.1). The Cons8 PCR product is shown in Figure 6.2.

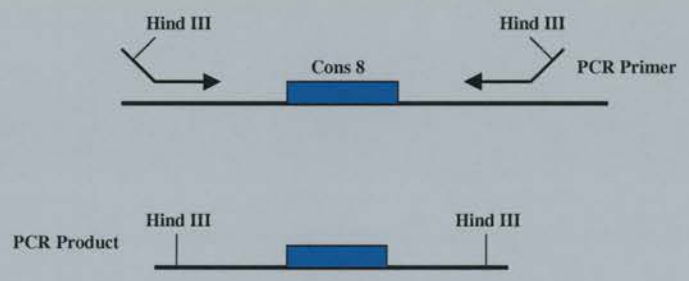


Fig 6.1: The PCR strategy used to clone Cons 8. Primers incorporating Hind III site were designed 450 bp either side of Cons 8. The resulting PCR product could then be easily ligated into any construct following Hind III digestion.

The Cons8 PCR product was then purified from an agarose gel using a Qiagen gel purification kit, and digested with Hind III, to give the Cons8 PCR product Hind III "sticky ends". A reporter construct pBGZ40 # 1230, that contained a *LacZ* reporter gene coupled to a β-globin minimal promoter, was also cut with Hind III. The Cons8 PCR product was then ligated into pBGZ40 # 1230 such that it lay next to the β-globin minimal promoter (see Fig 6.3). The ligation mix was eletroporated into *E.coli* XL-1 blue cells, and plasmid DNA generated from 20 of the resulting colonies. Diagnostic restriction digests were then carried out on plasmid DNA from each colony using NotI, NotI and SalI, and Hind III, to establish which clones contained pBGZ40 # 1230, with the desired Cons8 insert. On the basis of these

digest one clone (ML6) was selected for sequencing to ensure that its Cons8 insert had not picked up any mutations during the PCR amplification. The results of NotI, NotI and SalI, and Hind III digests on ML6 are shown in Fig. 6.2 and a construct map of ML6 is shown in Fig 6.3. The sequence of ML6's Cons8 insert proved to be identical to that of genomic DNA, i.e. no mutations had been introduced as a result of the PCR amplification process.

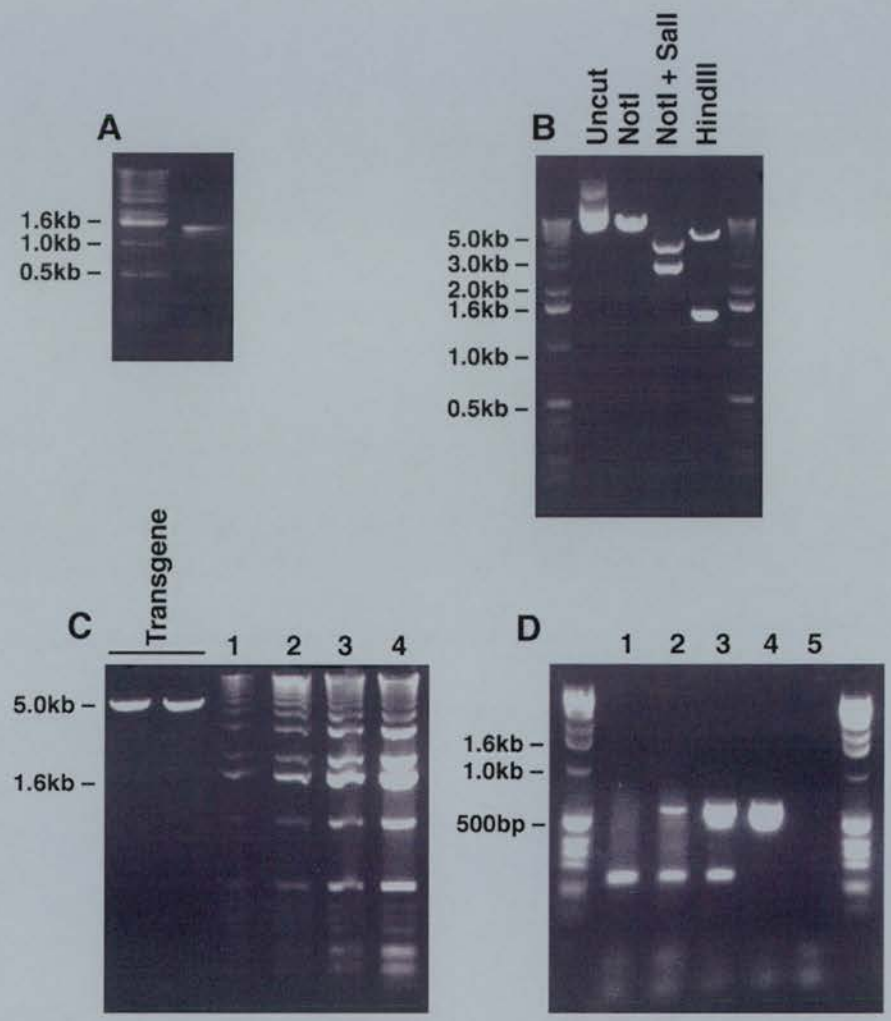


Fig 6.2: Gel pictures showing the key stages of the cloning of ML6 and the creation of the ML6 transgene. **Gel A:** the 1.5kb PCR product Cons 8 PCR product generated from mouse genomic DNA. **Gel B:** Diagnostic digests of ML6. **Gel C:** the concentration of ML6 transgene ready for injection is estimated by comparing it to DNA molecular weight markers (MW) of a known concentration. The 1.6 kb band is at the following concentration, 10ng/µl (lane 1), 20ng/µl (lane 2), 50ng/µl (lane 3) and 100ng/µl (lane 4). Both transgene lanes were estimated at 25ng/µl. **Gel D:** A sample of the PCRs carried out on embryo yolk sac DNA to establish which embryos carried the ML6 transgene. Lane 1, a non-transgenic embryo, lanes 2 and 3 show results from embryos that carry the transgene, lane 4 is a positive control for the ML6 transgene, lane 5 is a negative control. *LacZ* gives a band of 550 bp, and myosin a band 250bp.

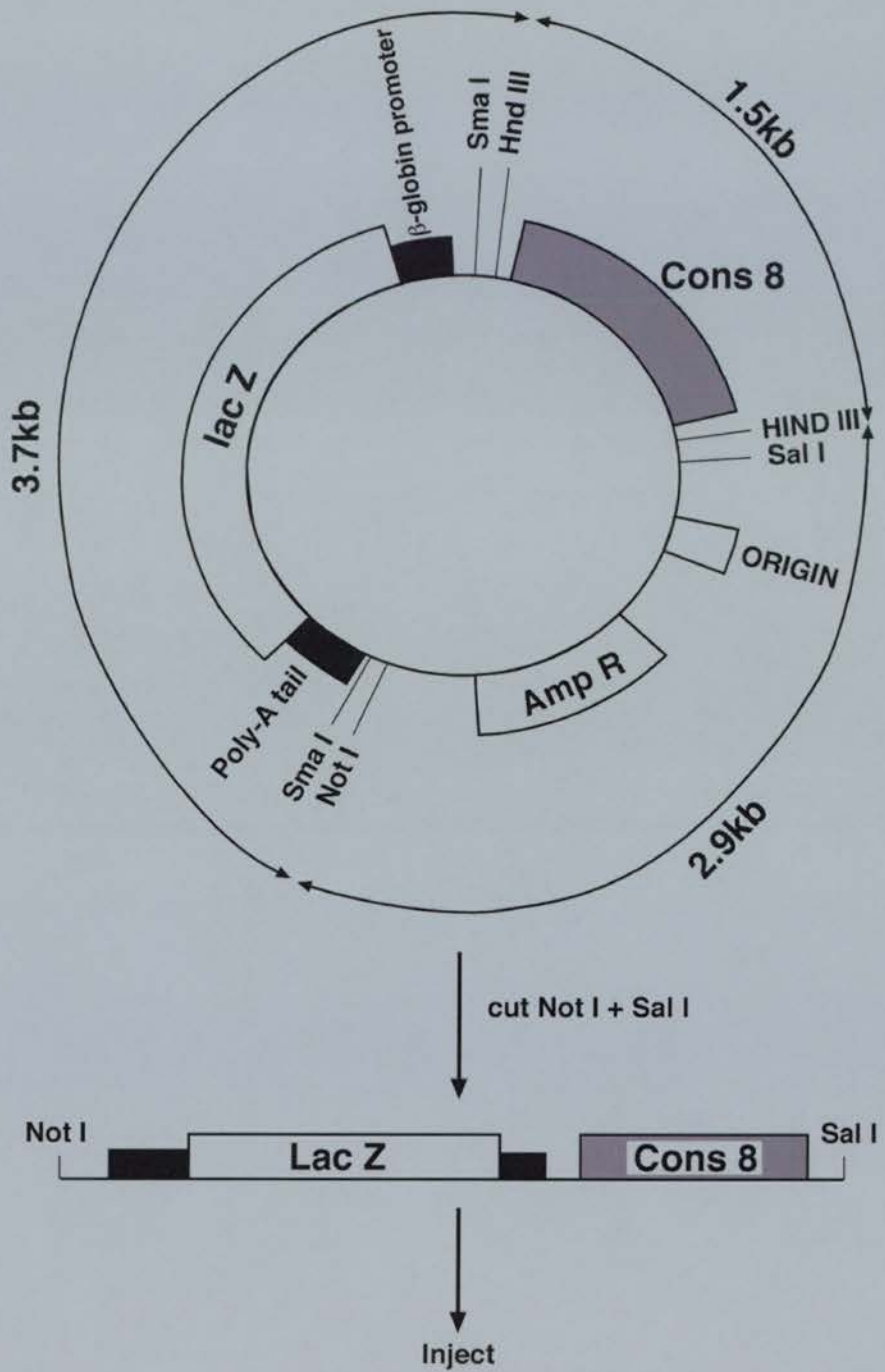


Fig 6.3: Diagrammatic representation of the construct ML6, showing restriction sites, and vector size. The linerised transgene is shown below the vector after digestion with Not I and Sal I.

6.2.2 Transgenic injection of ML6

To assess if Cons8 contained any enhancer sequences capable of driving gene expression in a limb specific pattern, it was necessary to prepare the construct for transgenic injection. This involved isolating the transgene, the Cons8 insert plus the β-globin mediated *LacZ* reporter, from the vector backbone. This was done by cutting ML6 with Not I and Sal I, to release the transgene, which was subsequently purified by gel extraction as mentioned in chapter 2. The concentration of the ML6 transgene was then estimated by comparison to DNA molecular weight markers of a known concentration, and injected into E0.5 mouse embryos at a concentration of 1ηg/μl. Injected embryos were transferred back into pseudo-pregnant females and allowed to develop to E10.5, whereupon they were harvested and stained for *LacZ* activity.

Yolk sacs were removed from harvested embryos and proteinase K digested to release genomic DNA, which was subsequently assessed for the presence of the transgene by PCR. The primers (P7 and P8) used to identify transgenic embryos were specific to the *LacZ* reporter gene of pBGZ40 # 1230, and generated a product of approximately 550bp. The *LacZ* primers were used in conjunction with a set of positive control primers (P9 and P10) specific to the gene *myosin* that created a PCR product of approximately 250bp. Three transgenic embryos were generated using ML6, (see Fig 6.2) but no *LacZ* staining was observed in any of the embryos (results not shown).

However, transgenic embryos generated by Dr. Laura Lettice and Mrs Lorna Marshall using an equivalent *LacZ* transgene but containing Cons9 rather that Cons8, did exhibit *LacZ* staining in the ZPA of limb buds, at E10.5 (see Fig 6.4). Additionally, a *LacZ* transgene incorporating Cons9 genomic sequence from *Fugu* was also able to drive *LacZ* expression, in a ZPA like pattern (see Fig 6.4). Thus Cons9 sequences from both mouse and *Fugu* were capable of driving gene expression in the ZPA of the limb, but the mouse Cons8 sequence did not exhibit any enhancer activity at E10.5.

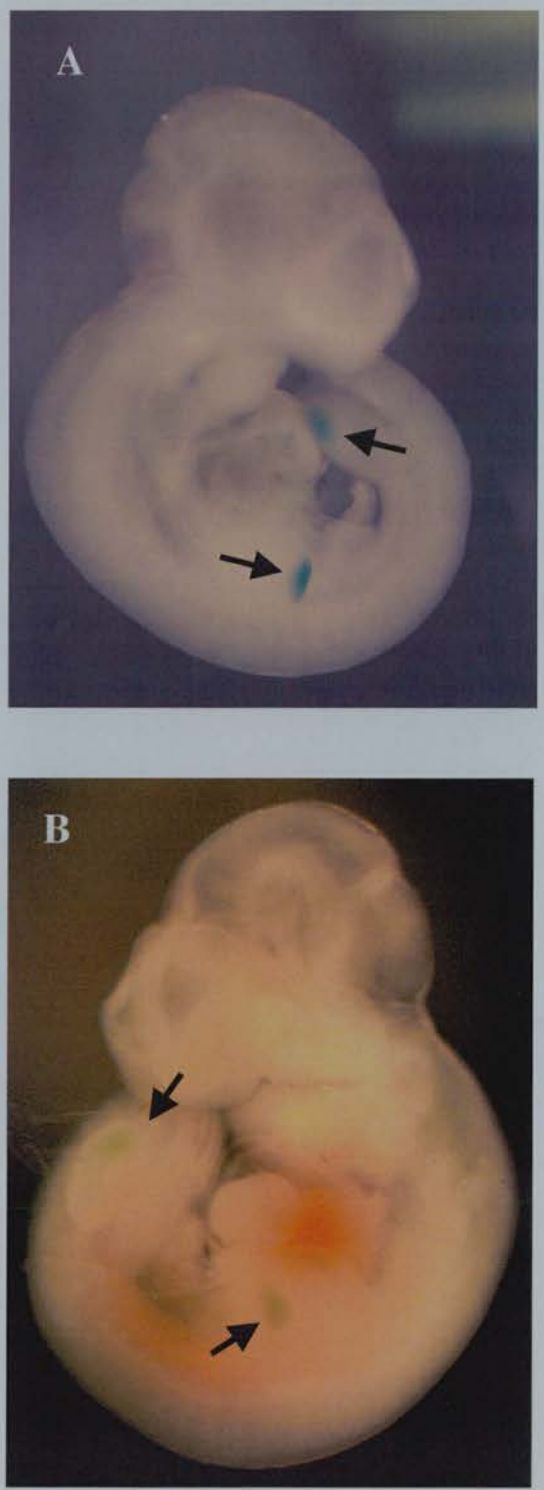


Fig 6.4: Transgenic E10.5 embryos created by Dr. Laura Lettice and Mrs Lorna Marshall. A: Transgenic embryo generated using mouse Cons 9 sequence linked to a *LacZ* reporter gene. B: Transgenic embryo generated using *Fugu* Cons 9 sequence linked to a *LacZ* reporter gene. Both embryos exhibit *LacZ* activity in the ZPA of limb buds, high-lighted by the black arrows.

6.3 Discussion

Enhancers were originally defined as *cis*-acting DNA sequences that drive gene transcription in a manner that is independent of their orientation and distance relative to the RNA start site (Edgar Serfling, 1985). Cons9 has been shown to be capable of driving gene expression within the ZPA of the limb bud. The closest gene to Cons9 that exhibits a ZPA expression pattern is *Shh*; therefore it seems most likely that Cons9 is responsible for driving transcription of *Shh* in the limb. Thus Cons9 seems to be capable of driving gene expression in *cis,* independent of distance, as it can act within several base pairs on *LacZ* or over nearly 1 Mb on *Shh*. Although it has not yet been confirmed that Cons9 can operate in an orientation independent manner, (this could be achieved by reversing the orientation of the Cons9 transgenic insert) it seems highly likely that Cons9 is a ZPA enhancer element of *Shh*, henceforth referred to as the *Ssq* enhancer. It also seems probable that *Ssq* enhancer is responsible for driving the expression of *HPAP* in the ZPA of *Ssq* limb buds, as Cons9 is positioned very close to the *Ssq* insertion site (they are both within *Lmbr1* intron 5).

The presence of *HPAP* staining in the limb buds of *Ssq* mice (see Chapter 1) was originally interpreted to mean that the *Ssq* mutation had revealed and disrupted *cis*-acting regulatory elements of *Shh* (1). The results of the *cis-trans* test in Chapter 3 confirmed that this was indeed the case. It now seems likely that the *Ssq* enhancer is the *cis*-acting regulatory element of *Shh* disrupted by the *Ssq* mutation, as not only does it exhibit enhancer activity consistent with the regulation of *Shh* in the limb. But the *Ssq* enhancer is also disrupted by the *Ssq* insertion event (the *Ssq* enhancer is within the genomic area duplicated by the *Ssq* mutation).

However, a great many questions still remain. Firstly it is unclear how a duplication of an enhancer such as Cons9 could result in the mis-regulation of *Shh* in the anterior of the limb. Additionally it is unknown how the other human and mouse limb mutations mapped close to *Ssq* interact with Cons9, or any of the other CNSs postulated to play a role in limb development. Finally, the function of Cons8, still remains a mystery. No enhancer like

activity was detected in the transgenic assays performed, though Cons8 was only assayed for enhancer activity at E10.5, leaving the possibility that it may have an enhancer function earlier or later in development. An alternative hypothesis is that Cons8 may have a gene regulatory role other than that of an enhancer. It has already been speculated in Chapter 3 that the control of *Shh* expression in the limb may well require the activity of a repressor, to restrict its expression to the ZPA. A good candidate for this respessor is Cons8, as it is a highly conserved non-coding sequence and in the right genomic region. A further possibility is that Cons8 might have no function, though this would seem unusual due to its high conservation between mammals and fish. A more thorough discussion incorporating the transgenic results with the rest of the data presented in this thesis is contained in chapter 7.

Chapter 7

Discussion

7.1 Discussion

Part of the aim of this thesis was to answer the question "Is the *Ssq* phenotype the result of the disruption of *Lmbr1,* or *cis*-acting regulatory elements of *Shh*"? Based on the *cis-trans* test in Chapter 3 and supporting evidence in Chapter 4, I can conclude that the *Ssq* phenotype is due to the disruption of *Shh cis*-acting limb regulatory elements. Further work in Chapters 5 and 6 addressed the second aim of the thesis by identifying several candidate genomic regions as potential regulatory elements; one of which was shown to act as an enhancer. However the identification of the *Ssq* mutation as a regulatory mutation of *Shh*, raises some interesting questions.

7.1.1 Finding the promoter

Perhaps the most immediate question, is how can the *Ssq* enhancer act to promote gene expression over the 800kb that separates it from its target, the *Shh* promoter? Enhancers are believed to act by facilitating the formation of large transcriptional complexes at gene promoters, from which transcription can be readily initiated and reinitiated. As such enhancer sequences are believed to contain DNA sequences that bind transcription factors, which are then brought into the transcription complex at the promoter. The method by which this occurs is currently the subject of much debate, though one attractive concept concerns DNA looping (Rippe et al., 1995a). This theory suggests that enhancers bind transcription factors, which are brought into close contact with the promoter by DNA folding, such that the DNA between the enhancer and promoter is looped out (see Fig 7.1).

DNA looping seems a convincing explanation for *cis* interactions over relatively short distances, such as enhancers located only a few thousand base pairs distant from the transcription start point. Linking enhancer proteins to proteins of the transcription complex can be envisaged as increasing the local concentrations of the proteins, facilitating their interaction. But enhancers located tens or hundreds of thousands of base pairs away from their targets present a more serious challenge to the DNA looping model. At distances of more than 5000 base pairs it has been estimated that mere DNA connectivity between two sites

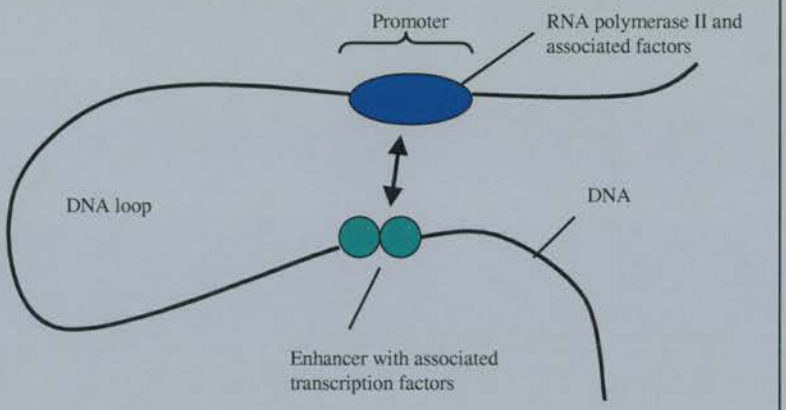


Fig 7.1: DNA looping is thought to bring proteins bound at distant enhancers to the promoter.

(promoter/enhancer) is not very effective in promoting protein-protein interactions in an equilibrium sense (Rippe et al., 1995b). In other words, sticking two proteins to opposite ends of a DNA molecule of a size greater than 5000 bp does not significantly increase the chance they will interact, compared to a DNA bound protein interacting with another protein free in solution (100-1000 molecules of a free protein per nucleus). Thus, it would appear that other mechanisms must act in conjunction with DNA looping to increase the chance that enhancer and promoter transcription factors interact.

Superhelicity in conjunction with nucleosome remodelling may help to bend and alter DNA structure in such a way as to facilitate enhancer-promoter interactions (Freeman and Garrard, 1992). Certainly it has been shown that packaging of the DNA into chromatin appears to promote long-distance interactions among DNA-bound factors (Barton et al., 1993). Raising the possibility that in addition to regulatory elements such as enhancers simply binding transcription factors, they may also be responsible for re-configuring chromatin structure to bend and loop DNA in such a way as to promote enhancer/promoter interactions. Though the affect of chromatin structure on transcription could of course be due to simple compaction of the DNA, thus reducing the distance between enhancer and promoter.

An alternative mechanism to DNA looping has been proposed and involves DNA scanning (Blackwood and Kadonaga, 1998). In a simple scanning model, enhancer-binding factors bind to their recognition sequences and then move continuously along the DNA until

they encounter their target promoter. However, a scanning mechanism cannot explain how an enhancer can activate transcription from a tailed hairpin (Plon and Wang, 1986), or how an enhancer on one chromosome can activate transcription from an allelic promoter on another paired chromosome, a phenomenon known as transvection (Wu, 1993). A more recent explanation for enhancer action combines elements of both the DNA-looping and scanning models, in a "facilitated tracking mechanism" (Blackwood and Kadonaga, 1998). In this model, an enhancer bound complex "tracks" via small steps along the chromatin, perhaps forming small loops along the way, until it reaches its target promoter, whereupon a large stable loop is formed. In this model it is envisaged that the chromatin over which the enhancer complex travels may provide cues, or be modified in such a way as to facilitate the formation of the final large DNA loop.

Currently evidence in support of models of enhancer action is limited, and a great deal more experimental work is required. However the identification of the *Ssq* enhancer and the fact that it can act over such a large distance will help to facilitate research in this field. If DNA looping is the method by which enhancers operate, then due to the distance involved a large DNA loop should occur when the *Ssq* enhancer is active. Such a large loop should be easier to detect than DNA loops from enhancers located closer to their cognate promoters.

7.1.2 Enhancer Specificity

In addition to interacting with distant promoters, enhancers must also make sure they are promoting transcription of the correct gene, i.e. they must show specificity. In the case of the *Ssq* enhancer this seems a particularly difficult problem to resolve, as there are a great many genes located closer to the enhancer than *Shh*, indeed *RNF32*, and *C7orf13* actually lie in-between *Shh* and the *Ssq* enhancer.

Two possible mechanisms have been suggested which might achieve enhancer-promoter selectivity. First, there could be specific interactions between enhancer-binding proteins and factors that interact with the promoter. The autoregulatory element 1 (AE1) enhancer in *Drosophilia* provides an example of preferential interactions between an enhancer

and a core promoter. In its endogenous context the AE1 enhancer is equidistant from *Sex combs reduced* (*Scr*) and *fushi tarazu* (*ftz*) promoters, but it selectively activates *ftz* expression. The *ftz* promoter contains a TATA box, but the *Scr* promoter does not. In transgenic constructs, the AE1 enhancer can activate transcription from a TATA-less promoter, but only in the absence of a TATA containing promoter (Ohtsuki et al., 1998). Thus components of the promoter can be important in forming, specific enhancer interactions.

A second mechanism that is believed to regulate enhancer activity involves the action of transcriptional boundary elements, best exemplified by the properties of the *Drosophilia gypsy* retrotransposon (Geyer, 1997). The *gypsy* element comprises multiple binding sites for the Suppressor of Hairy-wing (SuHw) protein, and when placed between an enhancer and its promoter, the *gypsy* element blocks the ability of the enhancer to promote transcription. But the intrinsic properties of the enhancer element remain unaffected as it can still stimulate transcription away from the *gypsy* element. Thus the *gypsy* element acts to form a boundary to the activity of a particular enhancer.

A possibility is that a combination of boundary elements and promoter specificity could be acting on the *Ssq* enhancer to restrict its ability to promote gene expression in *Shh* only. A boundary element between the *Ssq* enhancer and the *Lmbr1* promoter, could restrict the *Ssq* enhancer's activity towards *Shh*. Promoter specificity, could then account for the *Ssq* enhancers inability to promote gene expression of *RNF32* and *C7orf13* (See Fig 7.2).

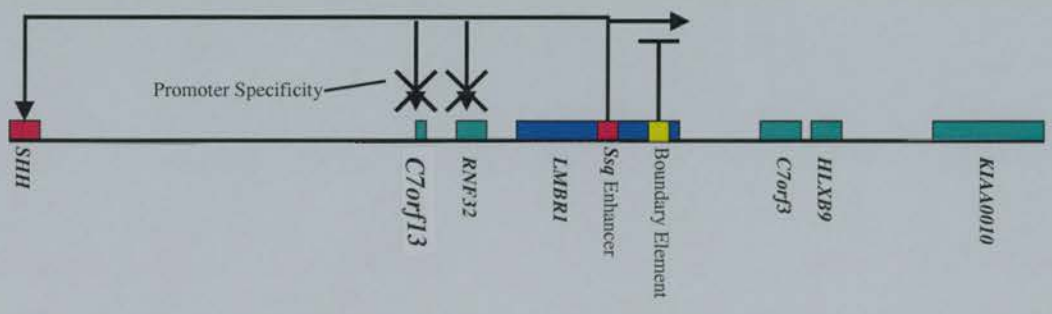


Fig 7.2: A possible scenario that enables the *Ssq* enhancer to act specifically on *SHH*. A boundary element in yellow inhibits the *Ssq* enhancer from acting on the genes to the right of *LMBR1*. Promoter specificity can account for the *Ssq* enhancers inability to stimulate gene expression from *RNF32* and *C7orf13*.

7.1.3 The Ssq insertion and the Ssq enhancer

As has been mentioned in chapter 6, Cons9 (the *Ssq* enhancer) is disrupted by the *Ssq* mutation, insofar that it is within the genomic area duplicated by the *Ssq* insertion event. It has been presumed that this duplication of the *Ssq* enhancer results in the anterior mis-expression of *Shh* within the limb bud, and therefore in the *Ssq* phenotype. However, it remains unclear how the duplication of an enhancer could result in the specific mis-regulation of a gene to the anterior of the limb. One possible explanation centres around the original hypothesis made in Chapter 3, that *Shh* must be actively repressed in the anterior of the limb bud by an anterior repressor, acting to inhibit the activity of the *Ssq* enhancer in the anterior limb bud. In *Ssq* mice it is possible to conceive a situation where there are more copies of the *Ssq* enhancer relative to the repressor, due to the duplication event at the *Ssq* insertion site. The anterior repressor may not be able to fully repress the extra copies of the *Ssq* enhancer in the anterior limb, facilitating the expression of *Shh* in the anterior limb bud.

7.1.4 The Ssq enhancer PPD, and Hx

Recently our collaborator Dr. Esther de Graaff (Dept. Clinical Genetics Erasmus University) has examined the *Ssq* enhancer sequence in seven families that exhibit limb specific PPD. In three of these families she has found single base pair changes in the CNS region of the *Ssq* enhancer, that segregate with affected individuals. These three point mutations are scattered over the *Ssq* enhancer, one mutation has affected a nucleotide conserved between mouse, human and *Fugu*, the other two alter a single base pair conserved between mouse and human but not *Fugu*. It is currently unclear if these single base changes are rare polymorphisms or the causative event for PPD. However their presence is suggestive that changes in the *Ssq* enhancer may result in an extra digit phenotype in humans as well as mice.

Dr. Esther de Graaff, has also found a point mutation in the *Hx* mouse which again is located within the CNS of the *Ssq* enhancer, affecting a base pair that is conserved between human, mouse and *Fugu*. The mutation segregates with the *Hx* phenotype, and is not present in the strain on which *Hx* was originally derived. Strongly suggesting that this single base pair

change is the causative event of the *Hx* mutation, and its presence once again implicates the *Ssq* enhancer in an extra-digit phenotype. The mechanism by which a single base pair change could affect the activity of the *Ssq* enhancer to promote anterior limb bud *Shh* expression is difficult to envisage. Though single base pair changes and small duplications (28bp) within other enhancer regions have been associated with subtle phenotypic effects in cystic fibrosis (Henry et al., 2001) and survival after chemotherapy treatment (Iacopetta et al., 2001). Thus it is possible that small changes in enhancer sequences may be a widespread disease mechanism, perhaps acting to effect changes within the protein binding sites of enhancers altering the transcriptional complexes that bind to them.

7.1.5 Multiple elements probably act in conjunction with the Ssq enhancer

Although the *Ssq* enhancer has been implicated in three PPD families and the *Hx* mouse, several PPD families are known to have normal *Ssq* enhancer sequences (Dr. Esther de Graaff personal communication). Additionally it is difficult to reconcile the Japanese translocation with the *Ssq* enhancer, as the translocation event is predicted to separate the *Ssq* enhancer from *Shh* (see Fig 5.15). This separation would be predicted to down-regulate *Shh* expression in the ZPA, rather than promote ectopic expression in the anterior limb bud, which is more consistent with a PPD phenotype. Thus although the *Ssq* enhancer is obviously important for *Shh* regulation in the limb, it seems likely that other regulatory elements are involved as well, such as the previously mentioned anterior repressor.

It is possible to envisage a scenario whereby mutations in an anterior repressor element could prevent the repression of *Shh* expression to the anterior limb bud, thus resulting in a PPD phenotype. This could then explain why some PPD families do not exhibit mutations in the *Ssq* enhancer, providing they demonstrate differences in an anterior repressor. However the Japanese translocation data still remains unresolved with this model, requiring the activity of a further regulatory element capable of driving anterior limb expression of *Shh* (a second enhancer), and remaining attached to *Shh* despite the translocation. Currently the anterior repressor and second enhancer elements remain speculations, but as mentioned in chapters 5,

several CNS regions with unknown function are located close to the *Ssq* enhancer, and are potential regulatory regions.

7.1.6 *Lmbr1* knockout and Acheiropodia

It was originally speculated in Chapter 3 that the Acheiropodia deletion may have removed a *Shh* enhancer, however no CNS regions were found in or near the genomic region deleted in Acheiropodia patients. Thus, currently it is difficult to attribute the cause of the Acheiropodia mutation to a disruption of *Shh* limb regulatory elements using comparative sequence analysis. However, recently Dr Laura Lettice and Dr Robert Watson (MRC, Human Genetics Unit, Edinburgh) have generated a mouse model of Acheiropodia. Using targeted ES cell recombination, a mouse was generated that lacked exon4 of *Lmbr1*, and most of the surrounding intronic DNA, but both heterozygotes and homozygotes for the mutation were wild type (Dr Laura Lettice personal communication).

The differences in the mouse and human Acheiropodia mutations are puzzling, but could be reconciled if there are differences in the location of a *Shh* limb regulatory element between the two species. A possibility is that a *Shh* regulatory element may have significantly shifted its position during the evolution of mice and humans, such that it is close to *LMBR1* exon 4 in humans but not so in mice. This could explain why a CNS region was not detected close to *LMBR1* exon 4 during the comparative sequence analysis, and the differences in mouse and humans exhibiting deletions around *Lmbr1* exon 4. However the *Lmbr1* transcript generated by the Acheiropodia mice is thought to be similar to the predicted *Lmbr1* transcript generated by Acheiropodia patients, i.e. to result in a severely truncated LMBR1 protein (Dr Laura Lettice personal communication), and probably acts as a *Lmbr1* null allele. The wild type phenotype of homozygous Acheiropodia mice, adds further doubt to the suggestion that *Lmbr1* acts to control *Shh* expression during development.

Obviously, the phenotype of Acheiropodia mice contradicts the partial *Lmbr1* null mice generated by Clark *et al.*, who have suggested that *Lmbr1* acts to control *Shh* expression in the limb. However, the two mutations have very different affects on *Lmbr1* transcription.

As described in Chapter 1, the *Lmbr1*^{ATG} allele reduces the level of transcription at the *Lmbr1* locus to approximately 7% of wild type levels. This dramatic reduction in transcription could have a pronounced affect on the chromatin structure of the *Lmbr1* locus, perhaps making it less accessible to transcription factors, such as those that bind to the *Ssq* enhancer. Acheiropodia mice, produce a truncated protein, but are not predicted to down-regulate *Lmbr1* transcription, probably resulting in a normal more open chromatin structure at the *Lmbr1* locus. Local chromatin structure is known to have a pronounced affect on the ability of enhancers to regulate gene expression, and enhancers have been found in introns of ubiquitously expressed genes (Iacopetta et al., 2001; Kleinjan et al., 2002). Suggesting that an open chromatin structure, such as that generated by continuously transcribed genes may be essential for some enhancers to function. This raises the intriguing possibility that the *Lmbr1*^{ATG} generated by Clark *et al.* may in fact be acting to down-regulate the activity of the *Ssq* enhancer by altering its local chromatin structure, to a closed less accessible form.

7.2 Conclusions and Further Work

The data presented in this thesis, provides direct evidence for an enhancer element acting over nearly 1Mb of genomic DNA, which is instrumental in the pathology of several mouse and human limb mutations. However we suspect that this is by no means a unique scenario within the vertebrate genome. Several other mutations have been attributed to long range position effects, which look like similar alterations in long-range gene regulatory elements (Iacopetta et al., 2001; Kleinjan et al., 2002; Belloni et al., 1996; Vortkamp et al., 1991; Wagner et al., 1994). However the discovery that the *Ssq* mutation is due to alterations in *cis*-acting gene regulatory elements of *Shh* raises more questions than it answers, some of which have been highlighted above. However some areas of further work to address these issues are outlined below.

A comprehensive transgenic assay of the remaining CNS regions identified in Chapter 5 could identify further enhancers of *Shh*. Cons 2-6 have already been speculated to contain enhancers that regulate *Shh* expression in the brain, and Cons 7, 8 and 11 may also

have some enhancer activity. Throughout this thesis, the speculation has been raised that *Shh* must be actively repressed in the anterior limb bud, by a repressor element. Cons 7, 8 and 11 are the most likely candidates for this repressor, and a thorough analysis of their sequence in PPD affected individuals would be desirable, particularly in those patients that do not exhibit point mutations in Cons 9. Assuming mutations were found in Cons 7, 8 or 11 that segregated with PPD, targeted recombination in ES cells could be used to generate mice carrying specific deletions in the appropriate Cons region. The resulting phenotype, could then help to establish if Cons 7, 8 or 11 have a function in *Shh* limb regulation. Extra digits could be interpreted to mean that an anterior repressor had been removed, where as a truncated limb phenotype would indicate that a further limb enhancer of *Shh* had been deleted.

The identification of the *Ssq* enhancer facilitates the identification of transcription factors that act to regulate *Shh* expression in the ZPA. A possible strategy would involve the use of a yeast-one-hybrid assay to identify transcription factors from a library, which could then be further assessed using ES cell recombination or transgenic assays. Further analysis of the *Ssq* enhancer sequence may also reveal known transcription factor binding sites.

References

Abi-Rached,L., Gilles,A., Shiina,T., Pontarotti,P., and Inoko,H. (2002). Evidence of en bloc duplication in vertebrate genomes. Nat. Genet. *31*, 100-105.

Ahn,D.G., Kourakis,M.J., Rohde,L.A., Silver,L.M., and Ho,R.K. (2002). T-box gene tbx5 is essential for formation of the pectoral limb bud. Nature *417*, 754-758.

Akimenko,M.A. and Ekker,M. (1995). Anterior duplication of the Sonic hedgehog expression pattern in the pectoral fin buds of zebrafish treated with retinoic acid. Dev. Biol *170*, 243-247.

Altabef,M., Clarke,J.D., and Tickle,C. (1997). Dorso-ventral ectodermal compartments and origin of apical ectodermal ridge in developing chick limb. Development *124*, 4547-4556.

Altabef,M. and Tickle,C. (2002). Initiation of dorso-ventral axis during chick limb development. Mech. Dev. *116*, 19-27.

Amores,A., Force,A., Yan,Y.L., Joly,L., Amemiya,C., Fritz,A., Ho,R.K., Langeland,J., Prince,V., Wang,Y.L., Westerfield,M., Ekker,M., and Postlethwait,J.H. (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711-1714.

Aparicio,S., Morrison,A., Gould,A., Gilthorpe,J., Chaudhuri,C., Rigby,P., Krumlauf,R., and Brenner,S. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. Proc. Natl. Acad. Sci. U. S. A *92*, 1684-1688.

Aza-Blanc,P., Ramirez-Weber,F.A., Laget,M.P., Schwartz,C., and Kornberg,T.B. (1997). Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. Cell *89*, 1043-1053.

Barton,M.C., Madani,N., and Emerson,B.M. (1993). The erythroid protein cGATA-1 functions with a stage-specific factor to activate transcription of chromatin-assembled beta-globin genes. Genes Dev. *7*, 1796-1809.

Belloni,E., Muenke,M., Roessler,E., Traverso,G., Siegel-Bartelt,J., Frumkin,A., Mitchell,H.F., Donis-Keller,H., Helms,C., Hing,A.V., Heng,H.H., Koop,B., Martindale,D., Rommens,J.M., Tsui,L.C., and Scherer,S.W. (1996). Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. Nat. Genet. *14*, 353-356.

Blackwood,E.M. and Kadonaga,J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 61-63.

Brenner,S., Elgar,G., Sandford,R., Macrae,A., Venkatesh,B., and Aparicio,S. (1993). Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature *366*, 265-268.

Bumcrot,D.A., Takada,R., and McMahon,A.P. (1995). Proteolytic processing yields two secreted forms of sonic hedgehog. Mol. Cell Biol *15*, 2294-2303.

Carrington,J.L. and Fallon,J.F. (1986). Experimental manipulation leading to induction of dorsal ectodermal ridges on normal limb buds results in a phenocopy of the Eudiplopodia chick mutant. Dev. Biol. *116*, 130-137.

Caruccio,N.C., Martinez-Lopez,A., Harris,M., Dvorak,L., Bitgood,J., Simandl,B.K., and Fallon,J.F. (1999). Constitutive activation of sonic hedgehog signaling in the chicken mutant talpid(2): Shh-independent outgrowth and polarizing activity. Dev. Biol *212*, 137-149.

Chang,D.T., Lopez,A., von Kessler,D.P., Chiang,C., Simandl,B.K., Zhao,R., Seldin,M.F., Fallon,J.F., and Beachy,P.A. (1994). Products, genetic linkage and limb patterning activity of a murine hedgehog gene. Development *120*, 3339-3353.

Charite,J., de Graaff,W., Shen,S., and Deschamps,J. (1994). Ectopic expression of Hoxb-8 causes duplication of the ZPA in the forelimb and homeotic transformation of axial structures. Cell *78*, 589-601.

Charite,J., McFadden,D.G., and Olson,E.N. (2000). The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development. Development *127*, 2461-2470.

Chen,Y. and Struhl,G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. Cell *87*, 553-563.

Chiang,C., Litingtung,Y., Harris,M.P., Simandl,B.K., Li,Y., Beachy,P.A., and Fallon,J.F. (2001). Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function. Dev. Biol *236*, 421-435.

Chiang,C., Litingtung,Y., Lee,E., Young,K.E., Corden,J.L., Westphal,H., and Beachy,P.A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature *383*, 407-413.

Chuang,P.T. and McMahon,A.P. (1999). Vertebrate Hedgehog signalling modulated by induction of a Hedgehog- binding protein. Nature *397*, 617-621.

Clark,R.M., Marker,P.C., and Kingsley,D.M. (2000). A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. Genomics *67*, 19-27.

Clark,R.M., Marker,P.C., Roessler,E., Dutra,A., Schimenti,J.C., Muenke,M., and Kingsley,D.M. (2001). Reciprocal mouse and human limb phenotypes caused by gain- and loss-of- function mutations affecting Lmbr1. Genetics *159*, 715-726.

Cohn,M.J. and Bright,P.E. (1999). Molecular control of vertebrate limb development, evolution and congenital malformations. Cell Tissue Res. *296*, 3-17.

Cohn,M.J., Izpisua-Belmonte,J.C., Abud,H., Heath,J.K., and Tickle,C. (1995). Fibroblast growth factors induce additional limb development from the flank of chick embryos. Cell *80*, 739-746.

Cohn,M.J., Patel,K., Krumlauf,R., Wilkinson,D.G., Clarke,J.D., and Tickle,C. (1997). Hox9 genes and vertebrate limb specification. Nature *387*, 97-101.

Cohn,M.J. and Tickle,C. (1999). Developmental basis of limblessness and axial patterning in snakes. Nature *399*, 474-479.

Dahn,R.D. and Fallon,J.F. (2000). Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. Science *289*, 438-441.

Davidson,D.R., Crawley,A., Hill,R.E., and Tickle,C. (1991). Position-dependent expression of two related homeobox genes in developing vertebrate limbs. Nature *352*, 429-431.

Dealy,C.N., Seghatoleslami,M.R., Ferrari,D., and Kosher,R.A. (1997). FGF-stimulated outgrowth and proliferation of limb mesoderm is dependent on syndecan-3. Dev. Biol. *184*, 343-350.

Dudley,A.T., Ros,M.A., and Tabin,C.J. (2002). A re-examination of proximodistal patterning during vertebrate limb development. Nature *418*, 539-544.

Duprez,D.M., Kostakopoulou,K., Francis-West,P.H., Tickle,C., and Brickell,P.M. (1996). Activation of Fgf-4 and HoxD gene expression by BMP-2 expressing cells in the developing chick limb. Development *122*, 1821-1828.

Echelard,Y., Epstein,D.J., St Jacques,B., Shen,L., Mohler,J., McMahon,J.A., and McMahon,A.P. (1993b). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417-1430.

Echelard,Y., Epstein,D.J., St Jacques,B., Shen,L., Mohler,J., McMahon,J.A., and McMahon,A.P. (1993a). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417-1430.

Edgar Serfling,M.J.W.S. (1985). Enhancers and eukaryoitic gene transcription. Trends Genet. *1*, 224-230.

Emori,Y. and Saigo,K. (1993). Distinct expression of two Drosophila homologs of fibroblast growth factor receptors in imaginal discs. FEBS Lett. *332*, 111-114.

Epstein,D.J., Martinu,L., Michaud,J.L., Losos,K.M., Fan,C., and Joyner,A.L. (2000). Members of the bHLH-PAS family regulate Shh transcription in forebrain regions of the mouse CNS. Development *127*, 4701-4709.

Fallon,J.F., Lopez,A., Ros,M.A., Savage,M.P., Olwin,B.B., and Simandl,B.K. (1994). FGF-2: apical ectodermal ridge growth signal for chick limb development. Science *264*, 104-107.

Favier,B. and Dolle,P. (1997). Developmental functions of mammalian Hox genes. Mol. Hum. Reprod. *3*, 115-131.

Feldman,B., Poueymirou,W., Papaioannou,V.E., DeChiara,T.M., and Goldfarb,M. (1995). Requirement of FGF-4 for postimplantation mouse development. Science *267*, 246-249.

Fernandez-Teran,M., Piedra,M.E., Simandl,B.K., Fallon,J.F., and Ros,M.A. (1997). Limb initiation and development is normal in the absence of the mesonephros. Dev. Biol. *189*, 246-255.

Francis-West,P.H., Robertson,K.E., Ede,D.A., Rodriguez,C., Izpisua-Belmonte,J.C., Houston,B., Burt,D.W., Gribbin,C., Brickell,P.M., and Tickle,C. (1995). Expression of genes encoding bone morphogenetic proteins and sonic hedgehog in talpid (ta3) limb buds: their relationships in the signalling cascade involved in limb patterning. Dev. Dyn. *203*, 187-197.

Francis,P.H., Richardson,M.K., Brickell,P.M., and Tickle,C. (1994). Bone morphogenetic proteins and a signalling pathway that controls patterning in the developing chick limb. Development *120*, 209-218.

Freeman,L.A. and Garrard,W.T. (1992). DNA supercoiling in chromatin structure and gene expression. Crit Rev. Eukaryot. Gene Expr. *2*, 165-209.

Fuse,N., Hirose,S., and Hayashi,S. (1996). Determination of wing cell fate by the escargot and snail genes in Drosophila. Development *122*, 1059-1067.

Galceran,J., Farinas,I., Depew,M.J., Clevers,H., and Grosschedl,R. (1999). Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev. *13*, 709-717.

Geduspan,J.S. and MacCabe,J.A. (1987). The ectodermal control of mesodermal patterns of differentiation in the developing chick wing. Dev. Biol. *124*, 398-408.

Geduspan,J.S. and Solursh,M. (1992). A growth-promoting influence from the mesonephros during limb outgrowth. Dev. Biol. *151*, 242-250.

Geyer,P.K. (1997). The role of insulator elements in defining domains of gene expression. Curr. Opin. Genet. Dev. *7*, 242-248.

Gibson-Brown,J.J., Agulnik,S.I., Chapman,D.L., Alexiou,M., Garvey,N., Silver,L.M., and Papaioannou,V.E. (1996). Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech. Dev. *56*, 93-101.

Goetinck P.F. (1964). Studies on limb morphogenesis II. Experiments with the polydactylous mutant eudiplopdia. Dev. Biol. *10*, 71-91.

Graham,A. (2000). The evolution of the vertebrates--genes and development. Curr. Opin. Genet. Dev. *10*, 624-628.

Grau Y,C.C.S.P. (1984). Mutations and chromosomal rearrangements affecting the expression of snail, a gene involved in embryonic patterning in Drosophilia. Genetics *108*, 347-360.

Guha,U., Gomes,W., Kobayashi,T., Pestell,R., and Kessler,J. (2002). In Vivo Evidence That BMP Signaling Is Necessary for Apoptosis in the Mouse Limb. Dev. Biol *249*, 108.

Hagan,D.M., Ross,A.J., Strachan,T., Lynch,S.A., Ruiz-Perez,V., Wang,Y.M., Scambler,P., Custard,E., Reardon,W., Hassan,S., Nixon,P., Papapetrou,C., Winter,R.M., Edwards,Y., Morrison,K., Barrow,M., Cordier-Alex,M.P., Correia,P., Galvin-Parton,P.A., Gaskill,S., Gaskin,K.J., Garcia-Minaur,S., Gereige,R., Hayward,R., and Homfray,T. (2000). Mutation analysis and embryonic expression of the HLXB9 Currarino syndrome gene. Am. J. Hum. Genet. *66*, 1504-1515.

Hardison,R.C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet. *16*, 369-372.

Hardison,R.C., Oeltjen,J., and Miller,W. (1997). Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res. *7*, 959-966.

Harrison,K.A., Thaler,J., Pfaff,S.L., Gu,H., and Kehrl,J.H. (1999). Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat. Genet. *23*, 71-75.

Henry,M.T., Cave,S., Rendall,J., O'Connor,C.M., Morgan,K., FitzGerald,M.X., and Kalsheker,N. (2001). An alpha1-antitrypsin enhancer polymorphism is a genetic modifier of pulmonary outcome in cystic fibrosis. Eur. J. Hum. Genet. *9*, 273-278.

Heus,H.C., Hing,A., van Baren,M.J., Joosse,M., Breedveld,G.J., Wang,J.C., Burgess,A., Donnis-Keller,H., Berglund,C., Zguricas,J., Scherer,S.W., Rommens,J.M., Oostra,B.A., and Heutink,P. (1999). A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. Genomics *57*, 342-351.

Heutink,P., Zguricas,J., van Oosterhout,L., Breedveld,G.J., Testers,L., Sandkuijl,L.A., Snijders,P.J., Weissenbach,J., Lindhout,D., Hovius,S.E., and . (1994). The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. Nat. Genet. *6*, 287-292.

Holland,P.W. and Garcia-Fernandez,J. (1996). Hox genes and chordate evolution. Dev. Biol. *173*, 382-395.

Hui,C.C. and Joyner,A.L. (1993). A mouse model of greig cephalopolysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. Nat. Genet. *3*, 241-246.

Iacopetta,B., Grieu,F., Joseph,D., and Elsaleh,H. (2001). A polymorphism in the enhancer region of the thymidylate synthase promoter influences the survival of colorectal cancer patients treated with 5-fluorouracil. Br. J. Cancer *85*, 827-830.

Ianakiev,P., van Baren,M.J., Daly,M.J., Toledo,S.P., Cavalcanti,M.G., Neto,J.C., Silveira,E.L., Freire-Maia,A., Heutink,P., Kilpatrick,M.W., and Tsipouras,P. (2001). Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. Am. J. Hum. Genet. *68*, 38-45.

Ingham,P.W. and McMahon,A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev. *15*, 3059-3087.

Isaac,A., Cohn,M.J., Ashby,P., Ataliotis,P., Spicer,D.B., Cooke,J., and Tickle,C. (2000). FGF and genes encoding transcription factors in early limb specification. Mech. Dev. *93*, 41-48.

Isaac,A., Rodriguez-Esteban,C., Ryan,A., Altabef,M., Tsukui,T., Patel,K., Tickle,C., and Izpisua-Belmonte,J.C. (1998). Tbx genes and limb identity in chick embryo development. Development *125*, 1867-1875.

Johnson,R.L. and Tabin,C.J. (1997). Molecular models for vertebrate limb development. Cell *90*, 979-990.

Kanegae,Y., Tavares,A.T., Izpisua Belmonte,J.C., and Verma,I.M. (1998). Role of Rel/NF-kappaB transcription factors during the outgrowth of the vertebrate limb. Nature *392*, 611-614.

Kawakami,Y., Capdevila,J., Buscher,D., Itoh,T., Rodriguez,E.C., and Izpisua Belmonte,J.C. (2001). WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo. Cell *104*, 891-900.

Kim,J., Irvine,K.D., and Carroll,S.B. (1995). Cell recognition, signal induction, and symmetrical gene activation at the dorsal-ventral boundary of the developing Drosophila wing. Cell *82*, 795-802.

Kleinjan,D.A., Seawright,A., Elgar,G., and van,H., V (2002). Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. Mamm. Genome *13*, 102-107.

Kumar,S. and Hedges,S.B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917-920.

Laufer,E., Dahn,R., Orozco,O.E., Yeo,C.Y., Pisenti,J., Henrique,D., Abbott,U.K., Fallon,J.F., and Tabin,C. (1997). Expression of Radical fringe in limb-bud ectoderm regulates apical ectodermal ridge formation. Nature *386*, 366-373.

Lee,J.J., Ekker,S.C., von Kessler,D.P., Porter,J.A., Sun,B.I., and Beachy,P.A. (1994). Autoproteolysis in hedgehog protein biogenesis. Science *266*, 1528-1537.

Lettice,L.A., Horikoshi,T., Heaney,S.J., van Baren,M.J., van der Linde,H.C., Breedveld,G.J., Joosse,M., Akarsu,N., Oostra,B.A., Endo,N., Shibata,M., Suzuki,M., Takahashi,E., Shinka,T., Nakahori,Y., Ayusawa,D., Nakabayashi,K., Scherer,S.W., Heutink,P., Hill,R.E., and Noji,S. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. U. S. A *99*, 7548-7553.

Lettice L, Hecksher-Sorensen J, Hill R.E. (1999) The dominant hemimelia mutation uncouples epithelial-mesenchymal interactions and disrupts anterior mesenchyme formation in mouse hindlimbs. Development *126*, 4729-4736

Lewandoski,M., Sun,X., and Martin,G.R. (2000). Fgf8 signalling from the AER is essential for normal limb development. Nat. Genet. *26*, 460-463.

Lewis,P.M., Dunn,M.P., McMahon,J.A., Logan,M., Martin,J.F., St Jacques,B., and McMahon,A.P. (2001). Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. Cell *105*, 599-612.

Litingtung,Y., Dahn,R.D., Li,Y., Fallon,J.F., and Chiang,C. (2002). Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. Nature *418*, 979-983.

Logan,C., Hornbruch,A., Campbell,I., and Lumsden,A. (1997). The role of Engrailed in establishing the dorsoventral axis of the chick limb. Development *124*, 2317-2324.

Loomis,C.A., Harris,E., Michaud,J., Wurst,W., Hanks,M., and Joyner,A.L. (1996). The mouse Engrailed-1 gene and ventral limb patterning. Nature *382*, 360-363.

Loots,G.G., Locksley,R.M., Blankespoor,C.M., Wang,Z.E., Miller,W., Rubin,E.M., and Frazer,K.A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science *288*, 136-140.

Lopez-Martinez,A., Chang,D.T., Chiang,C., Porter,J.A., Ros,M.A., Simandl,B.K., Beachy,P.A., and Fallon,J.F. (1995). Limb-patterning activity and restricted posterior localization of the amino-terminal product of Sonic hedgehog cleavage. Curr. Biol *5*, 791-796.

Lu,H.C., Revelli,J.P., Goering,L., Thaller,C., and Eichele,G. (1997). Retinoid signaling is required for the establishment of a ZPA and for the expression of Hoxb-8, a mediator of ZPA formation. Development *124*, 1643-1651.

MacKenzie,A., Purdie,L., Davidson,D., Collinson,M., and Hill,R.E. (1997). Two enhancer domains control early aspects of the complex expression pattern of Msx1. Mech. Dev. *62*, 29-40.

Martin,G.R. (1998). The roles of FGFs in the early development of vertebrate limbs. Genes Dev. *12*, 1571-1586.

Masuya,H., Sagai,T., Moriwaki,K., and Shiroishi,T. (1997). Multigenic control of the localization of the zone of polarizing activity in limb morphogenesis in the mouse. Dev. Biol *182*, 42-51.

Masuya,H., Sagai,T., Wakana,S., Moriwaki,K., and Shiroishi,T. (1995). A duplicated zone of polarizing activity in polydactylous mouse mutants. Genes Dev. *9*, 1645-1653.

Mayor,C., Brudno,M., Schwartz,J.R., Poliakov,A., Rubin,E.M., Frazer,K.A., Pachter,L.S., and Dubchak,I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. *16*, 1046-1047.

McLysaght,A., Hokamp,K., and Wolfe,K.H. (2002). Extensive genomic duplication during early chordate evolution. Nat. Genet. *31*, 200-204.

Meinhardt,H. (1983). A boundary model for pattern formation in vertebrate limbs. J. Embryol. Exp. Morphol. *76*, 115-137.

Mercader,N., Leonardo,E., Azpiazu,N., Serrano,A., Morata,G., Martinez,C., and Torres,M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature *402*, 425-429.

Mercader,N., Leonardo,E., Piedra,M.E., Martinez,A., Ros,M.A., and Torres,M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. Development *127*, 3961-3970.

Methot,N. and Basler,K. (1999). Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. Cell *96*, 819-831.

Methot,N. and Basler,K. (2001). An absolute requirement for Cubitus interruptus in Hedgehog signaling. Development *128*, 733-742.

Meyer,A. and Schartl,M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr. Opin. Cell Biol. *11*, 699-704.

Meyers,E.N., Lewandoski,M., and Martin,G.R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat. Genet. *18*, 136-141.

Michaud,J.L., Lapointe,F., and Le Douarin,N.M. (1997). The dorsoventral polarity of the presumptive limb is determined by signals produced by the somites and by the lateral somatopleure. Development *124*, 1453-1463.

Moon,A.M., Boulet,A.M., and Capecchi,M.R. (2000). Normal limb development in conditional mutants of Fgf4. Development *127*, 989-996.

Moon,A.M. and Capecchi,M.R. (2000). Fgf8 is required for outgrowth and patterning of the limbs. Nat. Genet. *26*, 455-459.

Moran,J.L., Levorse,J.M., and Vogt,T.F. (1999). Limbs move beyond the radical fringe. Nature *399*, 742-743.

Muller,F., Chang,B., Albert,S., Fischer,N., Tora,L., and Strahle,U. (1999). Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. Development *126*, 2103-2116.

Nanni,L., Ming,J.E., Bocian,M., Steinhaus,K., Bianchi,D.W., Die-Smulders,C., Giannotti,A., Imaizumi,K., Jones,K.L., Campo,M.D., Martin,R.A., Meinecke,P., Pierpont,M.E., Robin,N.H., Young,I.D., Roessler,E., and Muenke,M. (1999). The mutational spectrum of the sonic hedgehog gene in holoprosencephaly: SHH mutations cause a significant proportion of autosomal dominant holoprosencephaly. Hum. Mol. Genet. *8*, 2479-2488.

Niswander,L. and Martin,G.R. (1993). FGF-4 regulates expression of Evx-1 in the developing mouse limb. Development *119*, 287-294.

Niswander,L., Tickle,C., Vogel,A., Booth,I., and Martin,G.R. (1993). FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. Cell *75*, 579-587.

Ohno S. (1970). Evolution by Gene Duplication. (Berlin: Springer).

Ohtsuki,S., Levine,M., and Cai,H.N. (1998). Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes Dev. *12*, 547-556.

Ohuchi,H., Nakagawa,T., Yamamoto,A., Araga,A., Ohata,T., Ishimaru,Y., Yoshioka,H., Kuwana,T., Nohno,T., Yamasaki,M., Itoh,N., and Noji,S. (1997). The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. Development *124*, 2235-2244.

Parr,B.A. and McMahon,A.P. (1995). Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. Nature *374*, 350-353.

Pizette,S., Abate-Shen,C., and Niswander,L. (2001). BMP controls proximodistal outgrowth, via induction of the apical ectodermal ridge, and dorsoventral patterning in the vertebrate limb. Development *128*, 4463-4474.

Plon,S.E. and Wang,J.C. (1986). Transcription of the human beta-globin gene is stimulated by an SV40 enhancer to which it is physically linked but topologically uncoupled. Cell *45*, 575-580.

Porter,J.A., Young,K.E., and Beachy,P.A. (1996). Cholesterol modification of hedgehog signaling proteins in animal development. Science *274*, 255-259.

Pribnow,D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc. Natl. Acad. Sci. U. S. A *72*, 784-788.

Qu,S., Tucker,S.C., Ehrlich,J.S., Levorse,J.M., Flaherty,L.A., Wisdom,R., and Vogt,T.F. (1998). Mutations in mouse Aristaless-like4 cause Strong's luxoid polydactyly. Development *125*, 2711-2721.

Rancourt,D.E., Tsuzuki,T., and Capecchi,M.R. (1995). Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation. Genes Dev. *9*, 108-122.

Riddle,R.D., Ensini,M., Nelson,C., Tsuchida,T., Jessell,T.M., and Tabin,C. (1995). Induction of the LIM homeobox gene Lmx1 by WNT7a establishes dorsoventral pattern in the vertebrate limb. Cell *83*, 631-640.

Riddle,R.D., Johnson,R.L., Laufer,E., and Tabin,C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401-1416.

Rippe,K., von Hippel,P.H., and Langowski,J. (1995b). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem. Sci. *20*, 500-506.

Rippe,K., von Hippel,P.H., and Langowski,J. (1995a). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem. Sci. *20*, 500-506.

Rodriguez-Esteban,C., Schwabe,J.W., De La,P.J., Foys,B., Eshelman,B., and Belmonte,J.C. (1997). Radical fringe positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. Nature *386*, 360-366.

Rodriguez-Esteban,C., Schwabe,J.W., Pena,J.D., Rincon-Limas,D.E., Magallon,J., Botas,J., and Belmonte,J.C. (1998). Lhx2, a vertebrate homologue of apterous, regulates vertebrate limb outgrowth. Development *125*, 3925-3934.

Roessler,E., Belloni,E., Gaudenz,K., Jay,P., Berta,P., Scherer,S.W., Tsui,L.C., and Muenke,M. (1996). Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. Nat. Genet. *14*, 357-360.

Roessler,E. and Muenke,M. (2001). Midline and laterality defects: left and right meet in the middle. Bioessays *23*, 888-900.

Ros,M.A., Sefton,M., and Nieto,M.A. (1997). Slug, a zinc finger gene previously implicated in the early patterning of the mesoderm and the neural crest, is also involved in chick limb development. Development *124*, 1821-1829.

Rowe,D.A., Cairns,J.M., and Fallon,J.F. (1982). Spatial and temporal patterns of cell death in limb bud mesoderm after apical ectodermal ridge removal. Dev. Biol. *93*, 83-91.

Rowe,D.A. and Fallon,J.F. (1982). The proximodistal determination of skeletal parts in the developing chick leg. J. Embryol. Exp. Morphol. *68*, 1-7.

Ruvinsky,I. and Gibson-Brown,J.J. (2000). Genetic and developmental bases of serial homology in vertebrate limb evolution. Development *127*, 5233-5244.

Sanz-Ezquerro,J.J. and Tickle,C. (2000). Autoregulation of Shh expression and Shh induction of cell death suggest a mechanism for modulating polarising activity during chick limb development. Development *127* , 4811-4823.

Saunders,J.W.Jr. (1948). The proximo-distal sequence of origin of th parts of the chick wing and the role of the ectoderm. J. Exp. Zool. *108*, 363-403.

Saunders,J.W.Jr. and Ruess C.R. (1974). Inductive and Axial Properties of prospective Wing-Bud Mesoderm in the Embryo. Embryo. Dev. Biol. *38*, 41-50.

Saunders,J.W.Jr.G.M.T. (1968). Ectoderm-mesoderm interaction in the orgins of wing symmetry. (Baltimore: Williams and Wilkins), pp. 78-97.

Schimenti,J.C., Libby,B.J., Bergstrom,R.A., Wilson,L.A., Naf,D., Tarantino,L.M., Alavizadeh,A., Lengeling,A., and Bucan,M. (2000). Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells. Genome Res. *10*, 1043-1050.

Schwabe,J.W., Rodriguez-Esteban,C., and Izpisua Belmonte,J.C. (1998). Limbs are moving: where are they going? Trends Genet. *14*, 229-235.

Schwartz,S., Zhang,Z., Frazer,K.A., Smit,A., Riemer,C., Bouck,J., Gibbs,R., Hardison,R., and Miller,W. (2000). PipMaker--a web server for aligning two genomic DNA sequences. Genome Res. *10*, 577-586.

Sefton,M., Sanchez,S., and Nieto,M.A. (1998). Conserved and divergent roles for members of the Snail family of transcription factors in the chick and mouse embryo. Development *125*, 3111-3121.

Sekine,K., Ohuchi,H., Fujiwara,M., Yamasaki,M., Yoshizawa,T., Sato,T., Yagishita,N., Matsui,D., Koga,Y., Itoh,N., and Kato,S. (1999). Fgf10 is essential for limb and lung formation. Nat. Genet. *21*, 138-141.

Sharpe,J., Lettice,L., Hecksher-Sorensen,J., Fox,M., Hill,R., and Krumlauf,R. (1999). Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. Curr. Biol. *9*, 97-100.

Smith,J.C., Tickle,C., and Wolpert,L. (1978). Attenuation of positional signalling in the chick limb by high doses of gamma-radiation. Nature *272*, 612-613.

Summerbell D.Lewis J.H.Wolpert L. (1974). Positional Information in Chick Limb Morhogenesis. Nature *244*, 492-496.

Sun,X., Mariani,F.V., and Martin,G.R. (2002). Functions of FGF signalling from the apical ectodermal ridge in limb development. Nature *418*, 501-508.

Takeuchi,J.K., Koshiba-Takeuchi,K., Matsumoto,K., Vogel-Hopker,A., Naitoh-Matsuo,M., Ogura,K., Takahashi,N., Yasuda,K., and Ogura,T. (1999). Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds. Nature *398*, 810-814.

Tanaka,M., Cohn,M.J., Ashby,P., Davey,M., Martin,P., and Tickle,C. (2000). Distribution of polarizing activity and potential for limb formation in mouse and chick embryos and possible relationships to polydactyly. Development *127*, 4011-4021.

Tanaka,M., Munsterberg,A., Anderson,W.G., Prescott,A.R., Hazon,N., and Tickle,C. (2002). Fin development in a cartilaginous fish and the origin of vertebrate limbs. Nature *416*, 527-531.

te,W.P., Fernandez-Teran,M., Ros,M.A., and Zeller,R. (2002). Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. Genes Dev. *16*, 421-426.

Tickle,C. (1981). The number of polarizing region cells required to specify additional digits in the developing chick wing. Nature *289*, 295-298.

Tickle,C., Summerbell,D., and Wolpert,L. (1975). Positional signalling and specification of digits in chick limb morphogenesis. Nature *254*, 199-202.

Vogel,A., Rodriguez,C., and Izpisua-Belmonte,J.C. (1996). Involvement of FGF-8 in initiation, outgrowth and patterning of the vertebrate limb. Development *122*, 1737-1750.

Vogel,A., Rodriguez,C., Warnken,W., and Izpisua Belmonte,J.C. (1995). Dorsal cell fate specified by chick Lmx1 during vertebrate limb development. Nature *378*, 716-720.

Vortkamp,A., Gessler,M., and Grzeschik,K.H. (1991). GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. Nature *352*, 539-540.

Wagner,T., Wirth,J., Meyer,J., Zabel,B., Held,M., Zimmer,J., Pasantes,J., Bricarelli,F.D., Keutel,J., Hustert,E., and . (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell *79*, 1111-1120.

Wang,B., Fallon,J.F., and Beachy,P.A. (2000). Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. Cell *100*, 423-434.

Wojnar,P., Lechner,M., Merschak,P., and Redl,B. (2001). Molecular cloning of a novel lipocalin-1 interacting human cell membrane receptor using phage display. J. Biol. Chem. *276*, 20206-20212.

Wolpert L.Macpherson I.Todd I. (1969). Positional information and the spatial pattern of cellular differentiation. Cell spreading and cell movement: an active or a passive process? J. Theor Biol *25*, 1-47.

Wu,C.T. (1993). Transvection, nuclear structure, and chromatin proteins. J. Cell Biol *120*, 587-590.

Xu,J., Liu,Z., and Ornitz,D.M. (2000). Temporal and spatial gradients of Fgf8 and Fgf17 regulate proliferation and differentiation of midline cerebellar structures. Development *127*, 1833-1843.

Yokouchi,Y., Sakiyama,J., Kameda,T., Iba,H., Suzuki,A., Ueno,N., and Kuroiwa,A. (1996). BMP-2/-4 mediate programmed cell death in chicken limb buds. Development *122*, 3725-3734.

Zeller,R. and Duboule,D. (1997). Dorso-ventral limb polarity and origin of the ridge: on the fringe of independence? Bioessays *19*, 541-546.

Zeng,X., Goetz,J.A., Suber,L.M., Scott,W.J., Jr., Schreiner,C.M., and Robbins,D.J. (2001). A freely diffusible form of Sonic hedgehog mediates long-range signalling. Nature *411*, 716-720.

Zguricas,J., Heus,H., Morales-Peralta,E., Breedveld,G., Kuyt,B., Mumcu,E.F., Bakker,W., Akarsu,N., Kay,S.P., Hovius,S.E., Heredero-Baute,L., Oostra,B.A., and Heutink,P. (1999). Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the gene responsible to a 1.9 cM region on chromosome 7q36. J. Med. Genet. *36*, 32-40.

Zou,H. and Niswander,L. (1996). Requirement for BMP signaling in interdigital apoptosis and scale formation. Science *272*, 738-741

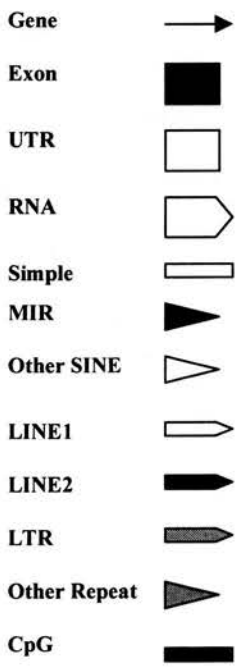
Errata:

The Pip and Vista plots in appendix 1and 2 have been mislabelled, *saph1* should read *Lmbr1*.

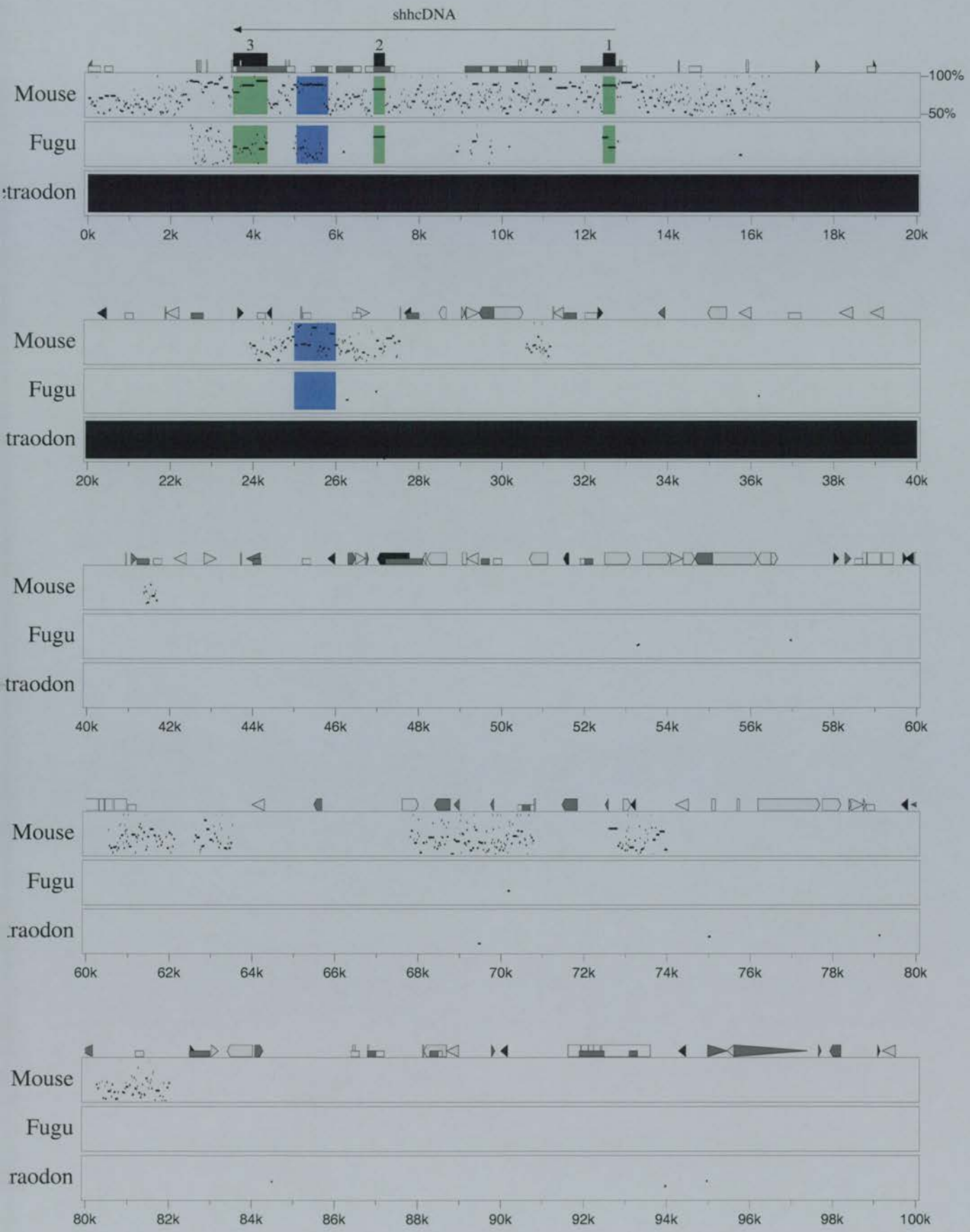
Appendix 1

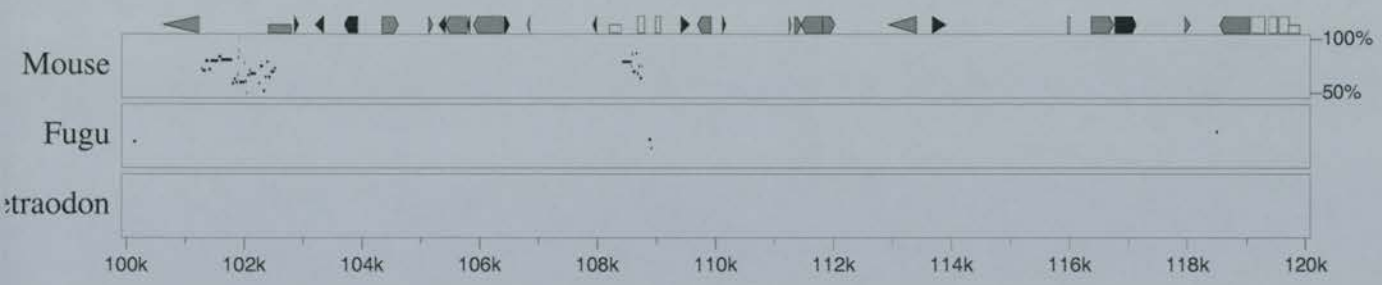
PIP Plots of genomic regions 1, 2 and 3

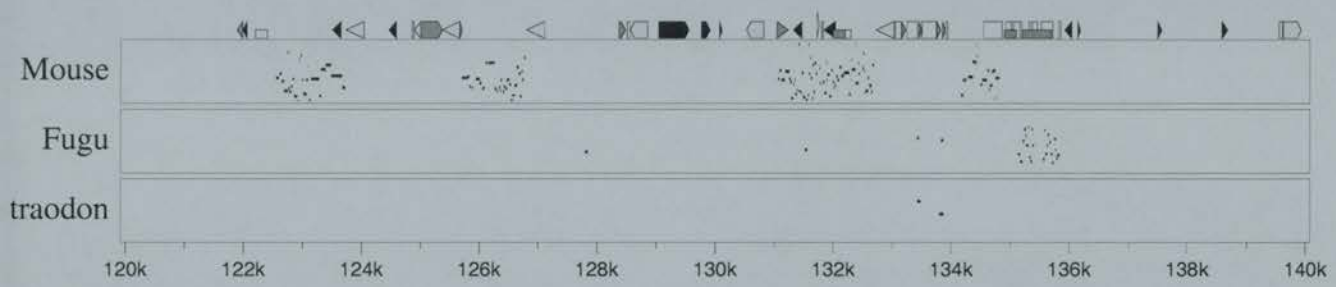
PIP Plot Key:

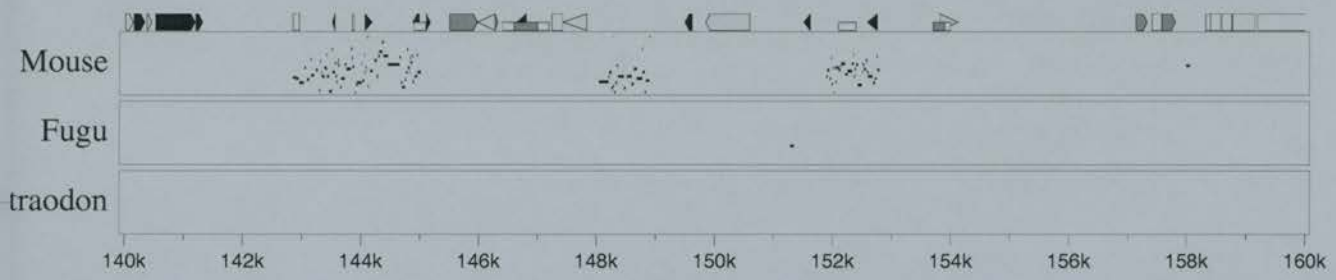


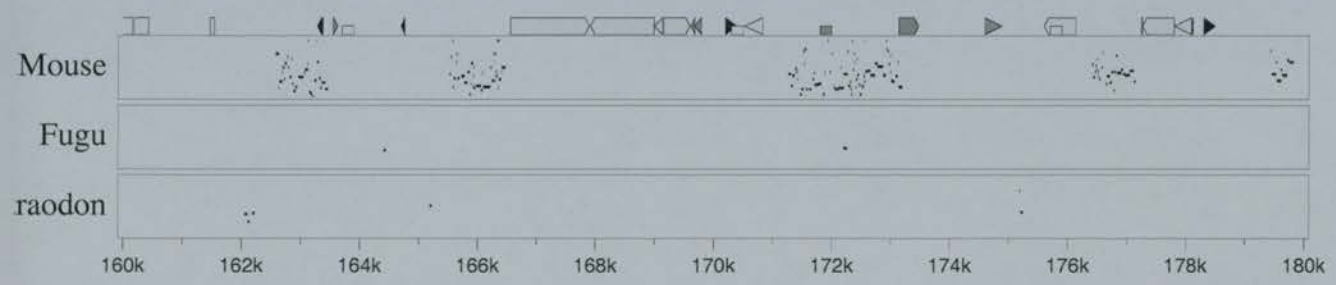
PIP plot *Shh* to *Lmbr1*

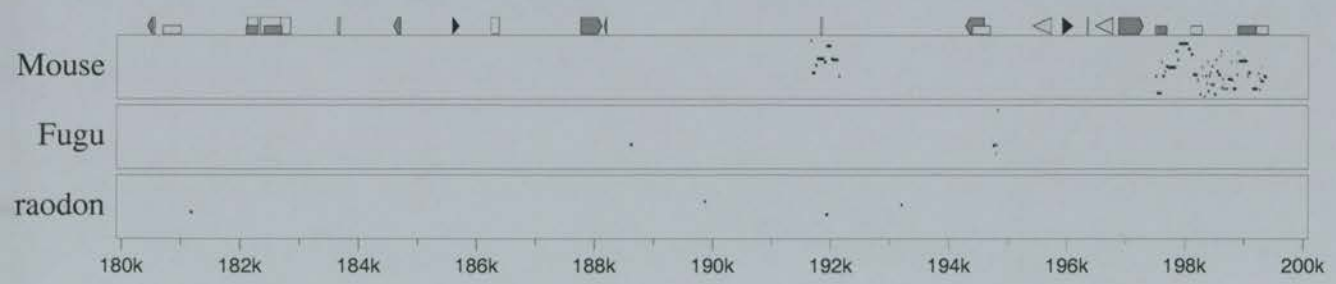

shhcDNA
3
2
1
Mouse
Fugu
traodon
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
Mouse
Fugu
traodon
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
Mouse
Fugu
traodon
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
Mouse
Fugu
raodon
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
Mouse
Fugu
raodon
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k

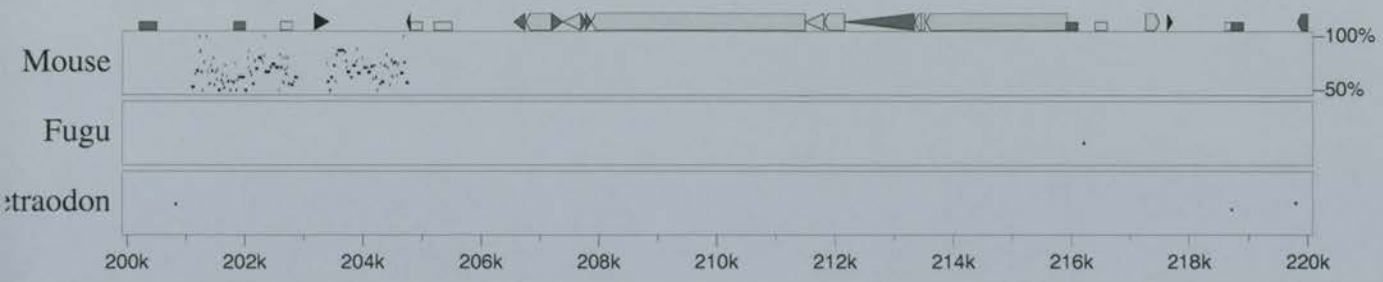

Mouse
Fugu
etraodon
100%
50%
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k

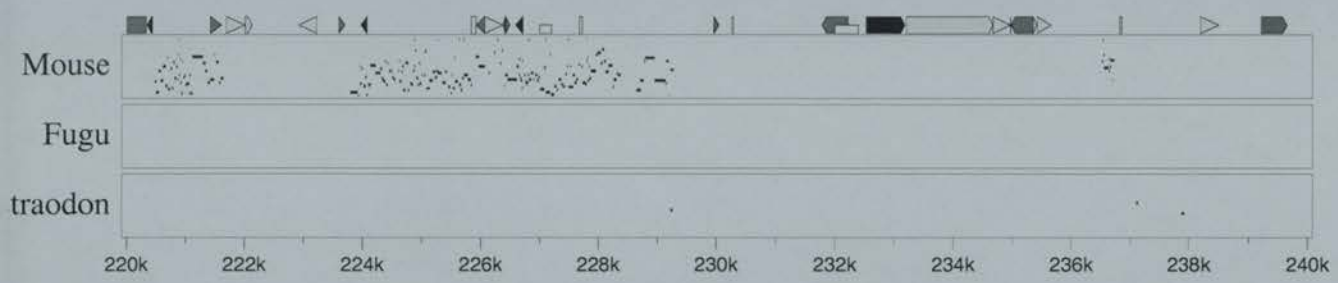

Mouse
Fugu
traodon
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k


Mouse
Fugu
traodon
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k

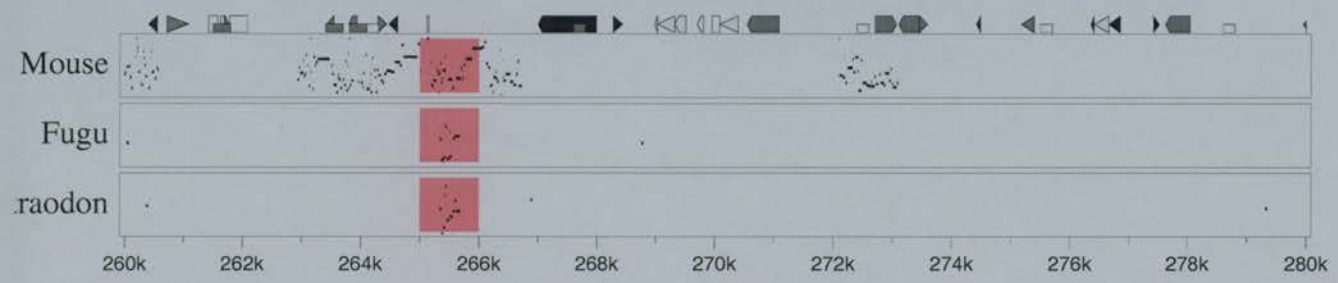

Mouse
Fugu
raodon
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k

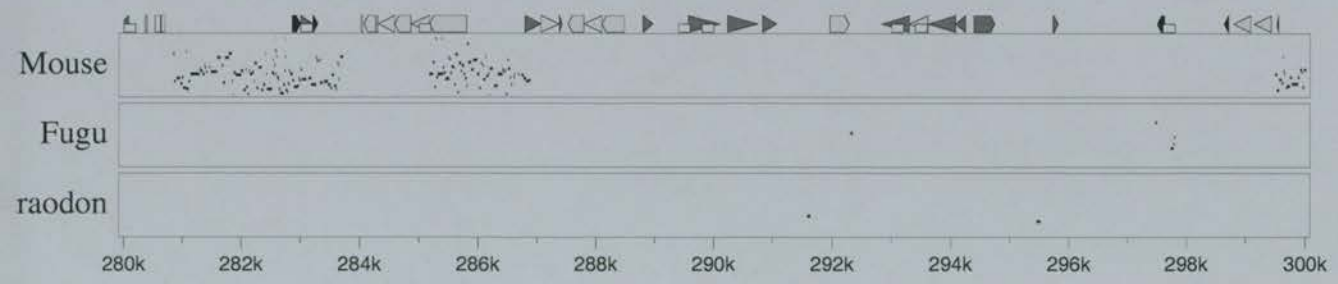

Mouse
Fugu
raodon
180k
182k
184k
186k
188k
190k
192k
194k
196k
198k
200k

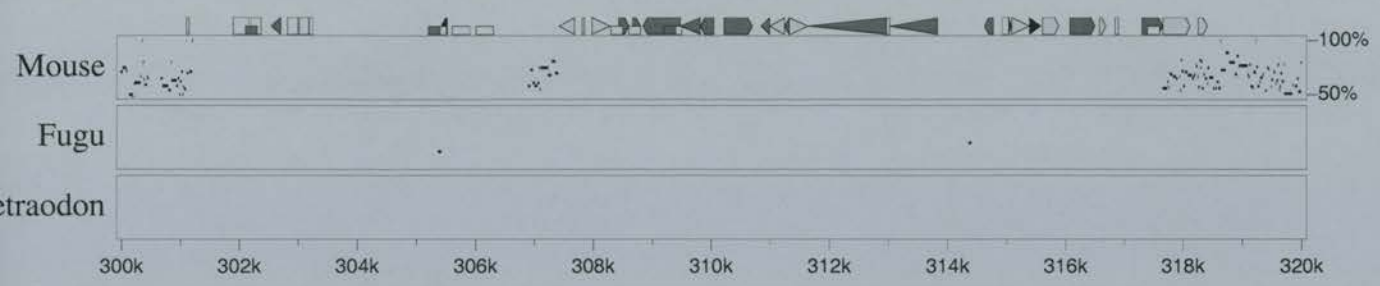

Mouse
Fugu
etraodon
100%
50%
200k
202k
204k
206k
208k
210k
212k
214k
216k
218k
220k

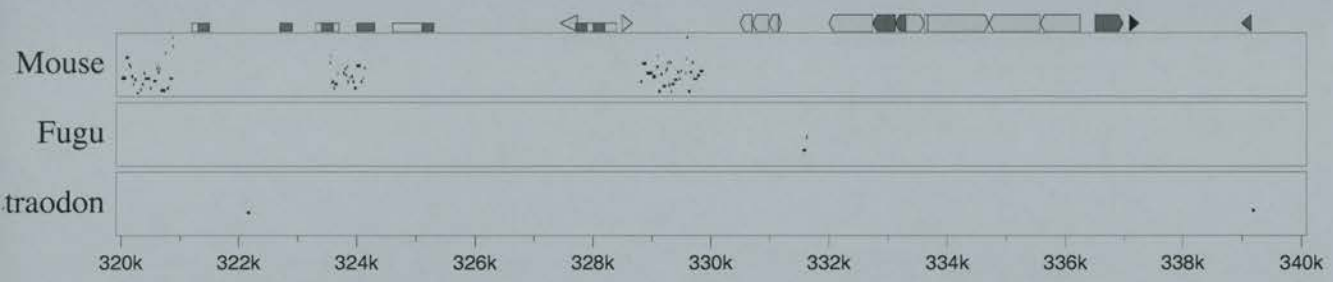

Mouse
Fugu
traodon
220k
222k
224k
226k
228k
230k
232k
234k
236k
238k
240k


Mouse
Fugu
traodon
240k
242k
244k
246k
248k
250k
252k
254k
256k
258k
260k

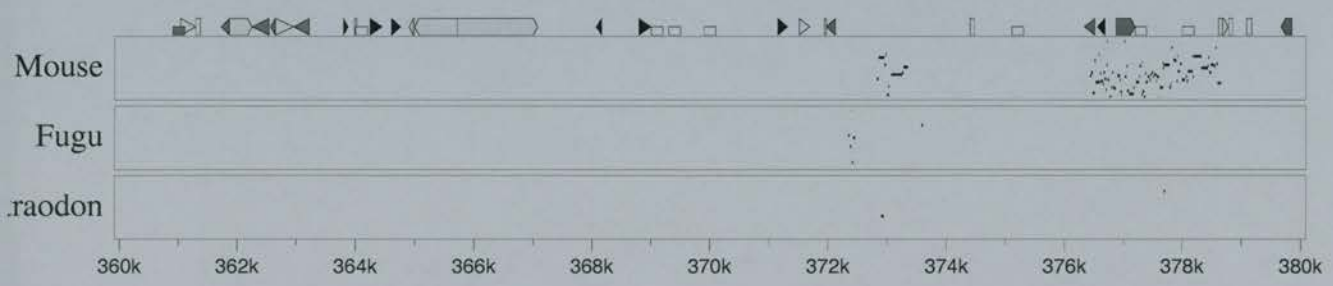

Mouse
Fugu
raodon
260k
262k
264k
266k
268k
270k
272k
274k
276k
278k
280k

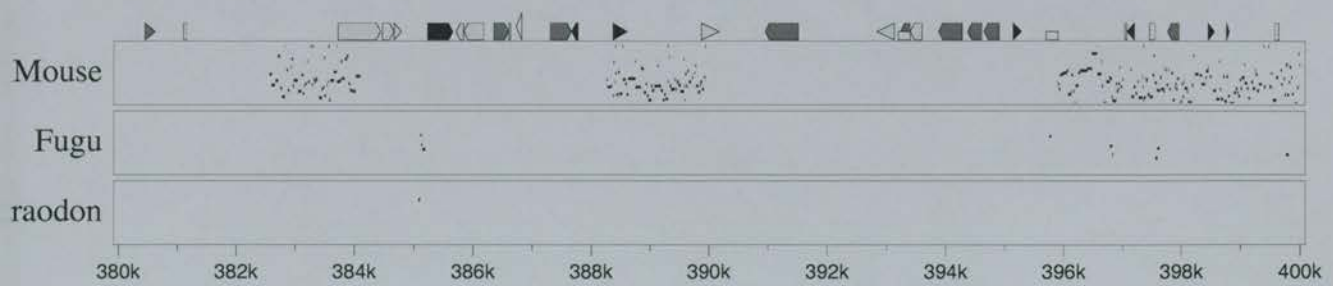

Mouse
Fugu
raodon
280k
282k
284k
286k
288k
290k
292k
294k
296k
298k
300k

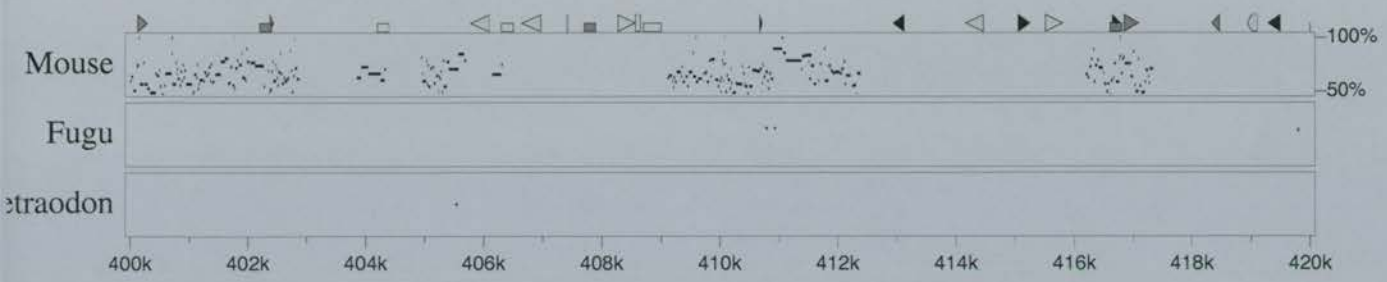

Mouse
Fugu
etraodon
100%
50%
300k
302k
304k
306k
308k
310k
312k
314k
316k
318k
320k

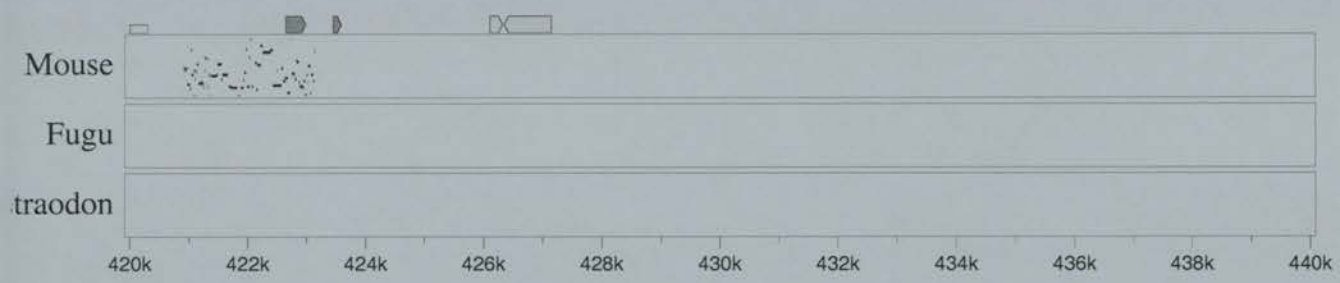

Mouse
Fugu
traodon
320k
322k
324k
326k
328k
330k
332k
334k
336k
338k
340k

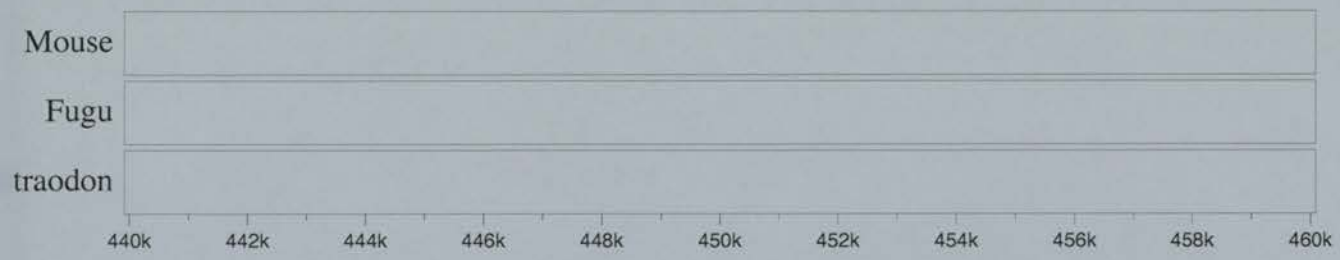

Mouse
Fugu
traodon
340k
342k
344k
346k
348k
350k
352k
354k
356k
358k
360k

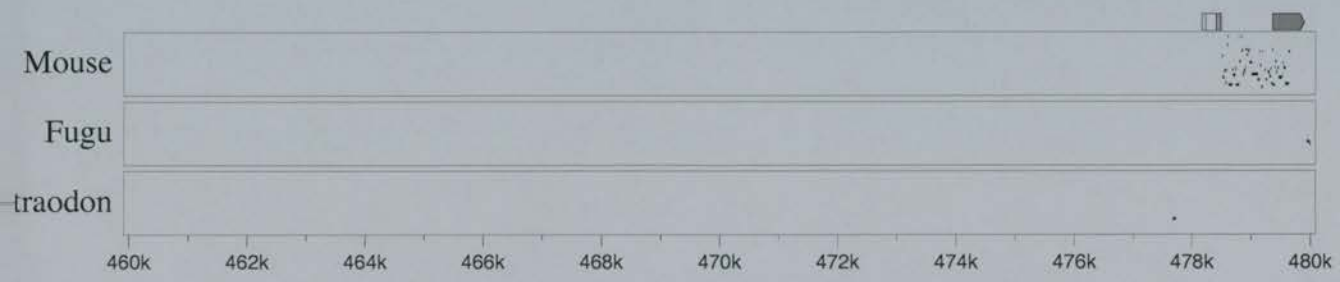

Mouse
Fugu
raodon
360k
362k
364k
366k
368k
370k
372k
374k
376k
378k
380k

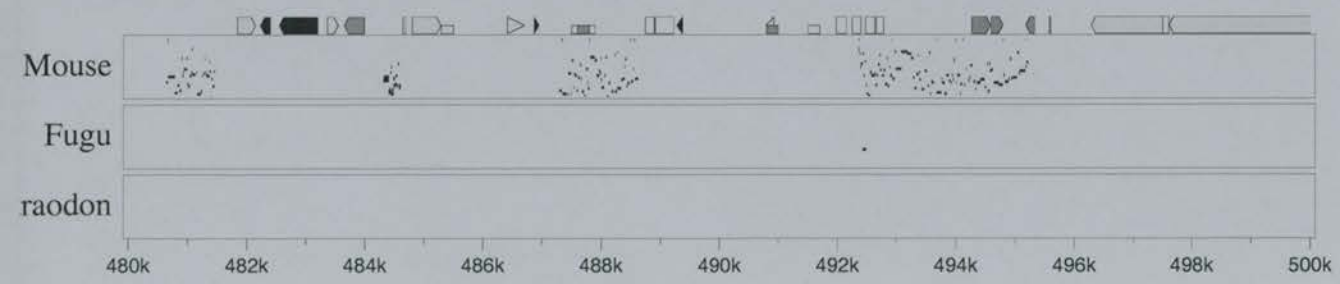

Mouse
Fugu
raodon
380k
382k
384k
386k
388k
390k
392k
394k
396k
398k
400k


Mouse
Fugu
etraodon
100%
50%
400k
402k
404k
406k
408k
410k
412k
414k
416k
418k
420k

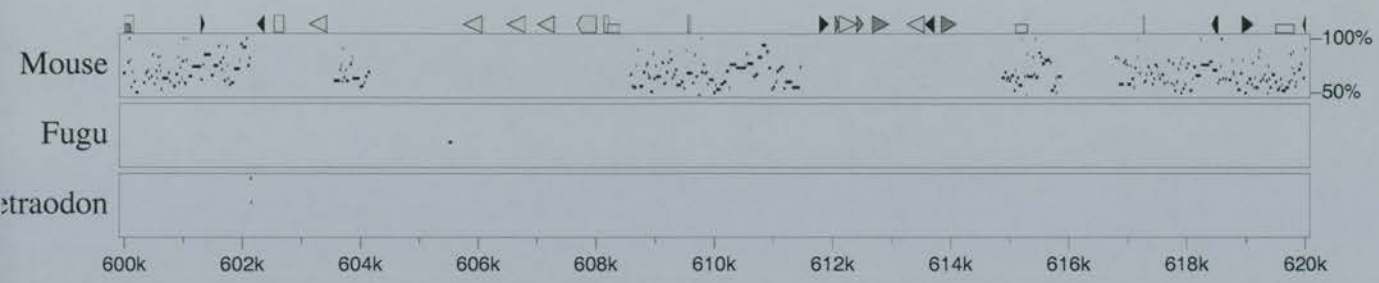

Mouse
Fugu
traodon
420k
422k
424k
426k
428k
430k
432k
434k
436k
438k
440k

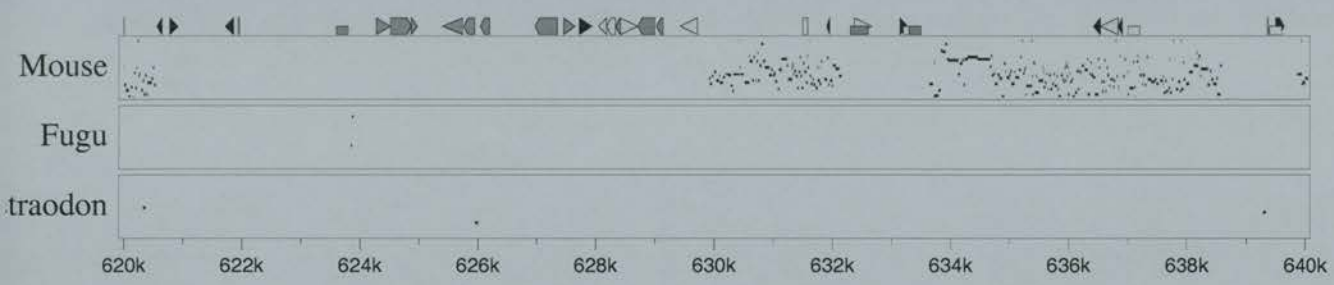

Mouse
Fugu
traodon
440k
442k
444k
446k
448k
450k
452k
454k
456k
458k
460k


Mouse
Fugu
traodon
460k
462k
464k
466k
468k
470k
472k
474k
476k
478k
480k

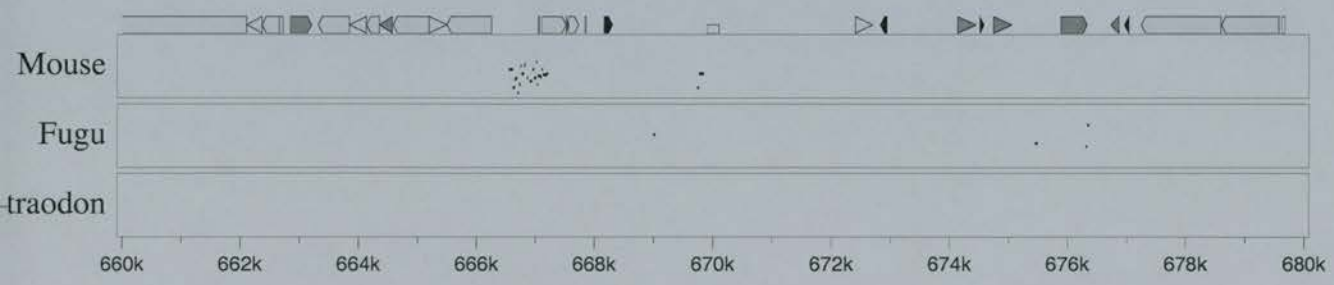

Mouse
Fugu
raodon
480k
482k
484k
486k
488k
490k
492k
494k
496k
498k
500k

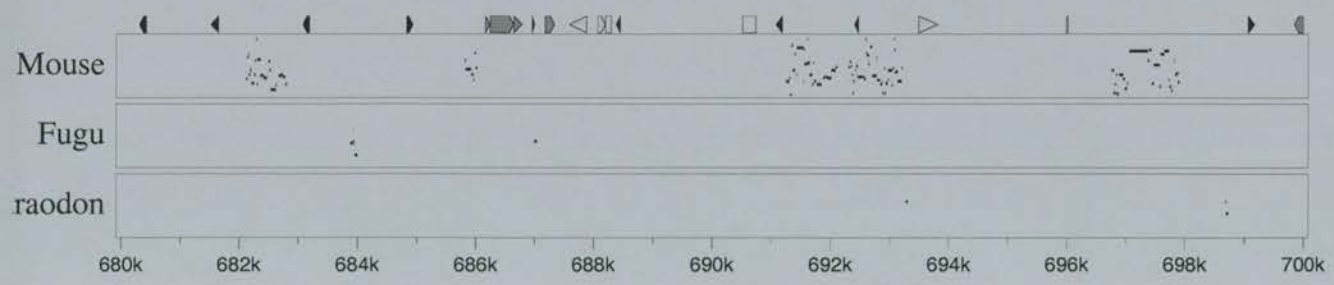

Mouse
Fugu
etraodon
100%
50%
500k 502k 504k 506k 508k 510k 512k 514k 516k 518k 520k
Mouse
Fugu
etraodon
520k 522k 524k 526k 528k 530k 532k 534k 536k 538k 540k
Mouse
Fugu
traodon
540k 542k 544k 546k 548k 550k 552k 554k 556k 558k 560k
Mouse
Fugu
traodon
560k 562k 564k 566k 568k 570k 572k 574k 576k 578k 580k
Mouse
Fugu
raodon
580k 582k 584k 586k 588k 590k 592k 594k 596k 598k 600k

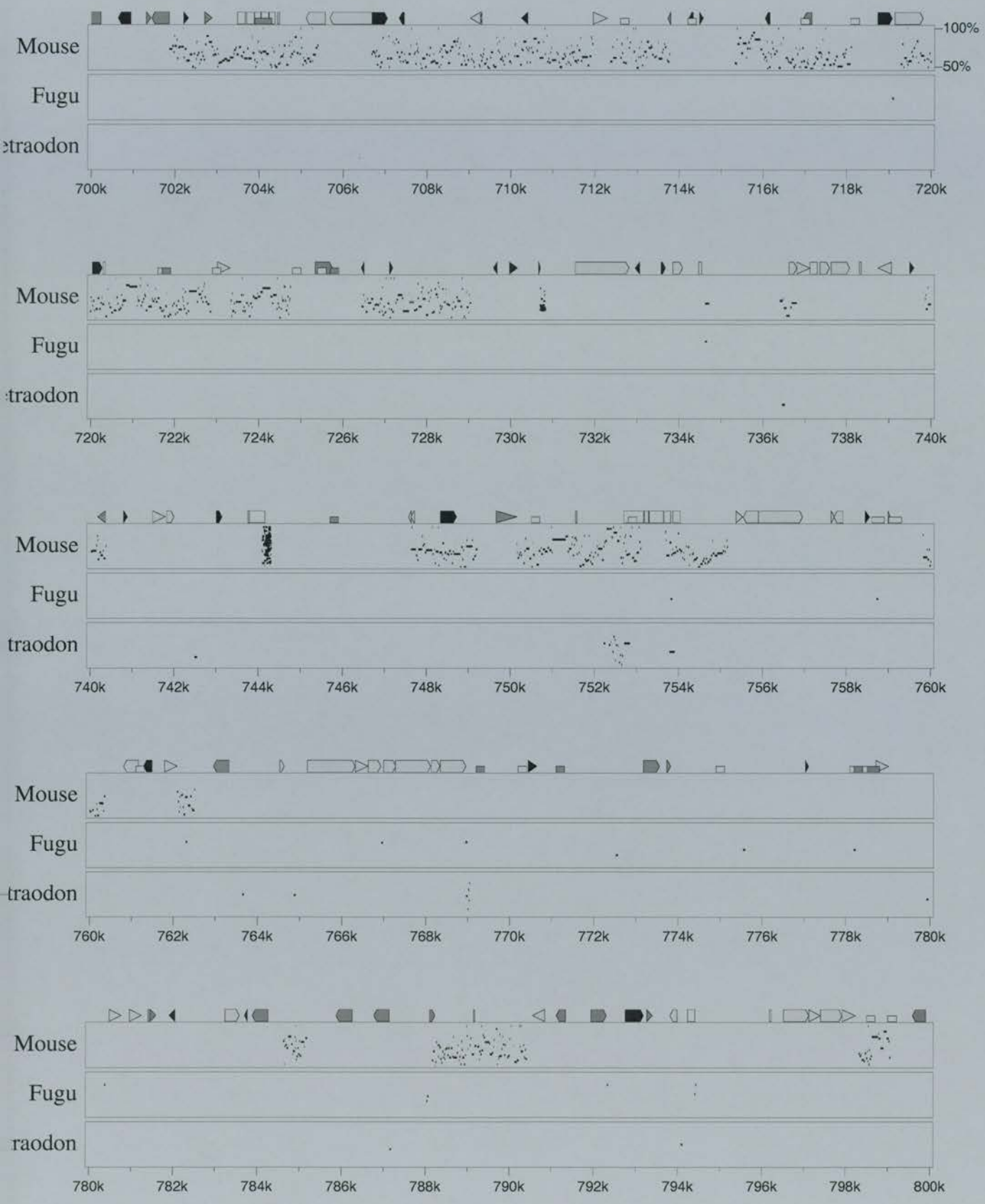

Mouse
Fugu
etraodon
100%
50%
600k
602k
604k
606k
608k
610k
612k
614k
616k
618k
620k


Mouse
Fugu
traodon
620k
622k
624k
626k
628k
630k
632k
634k
636k
638k
640k

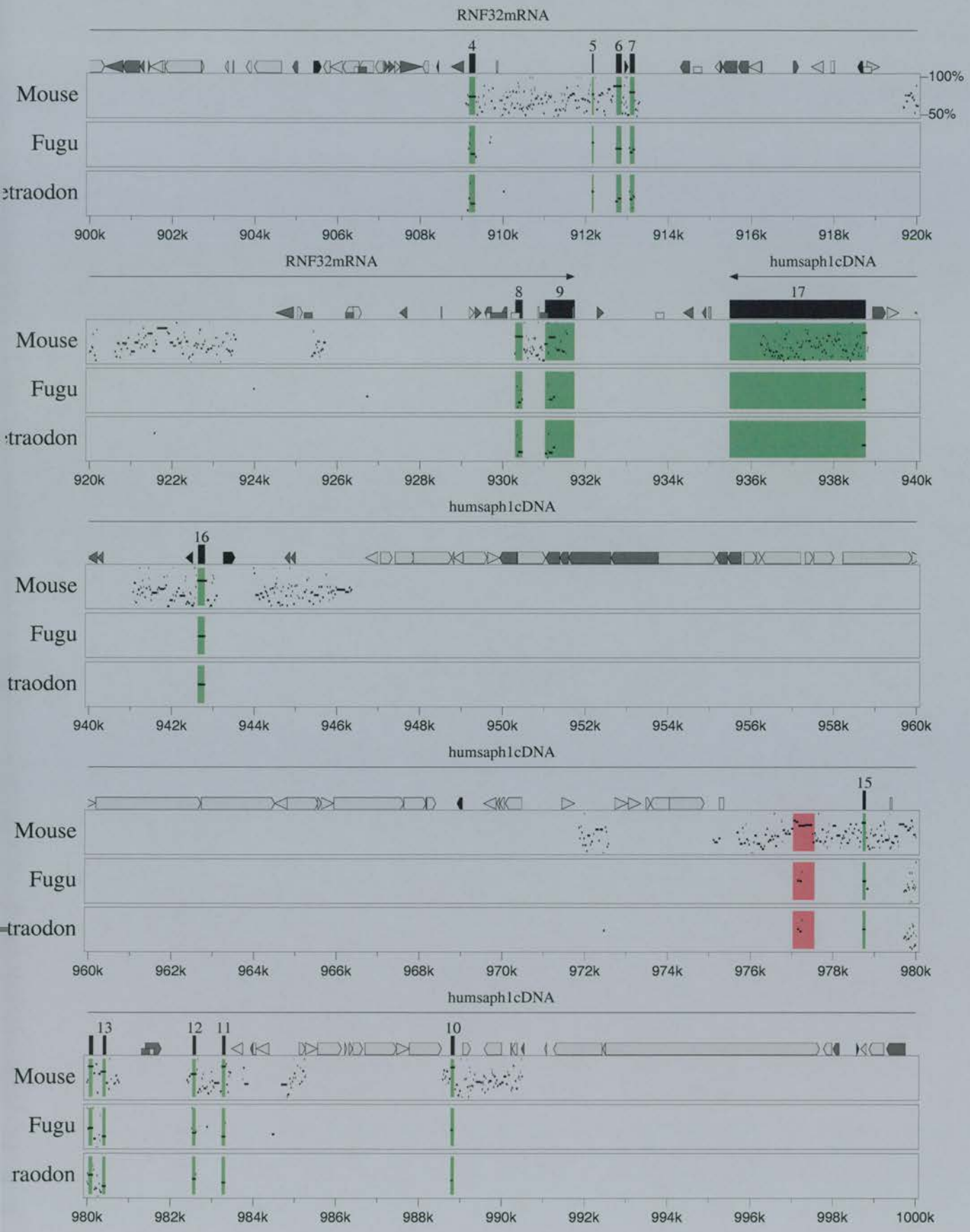

Mouse
Fugu
traodon
640k
642k
644k
646k
648k
650k
652k
654k
656k
658k
660k

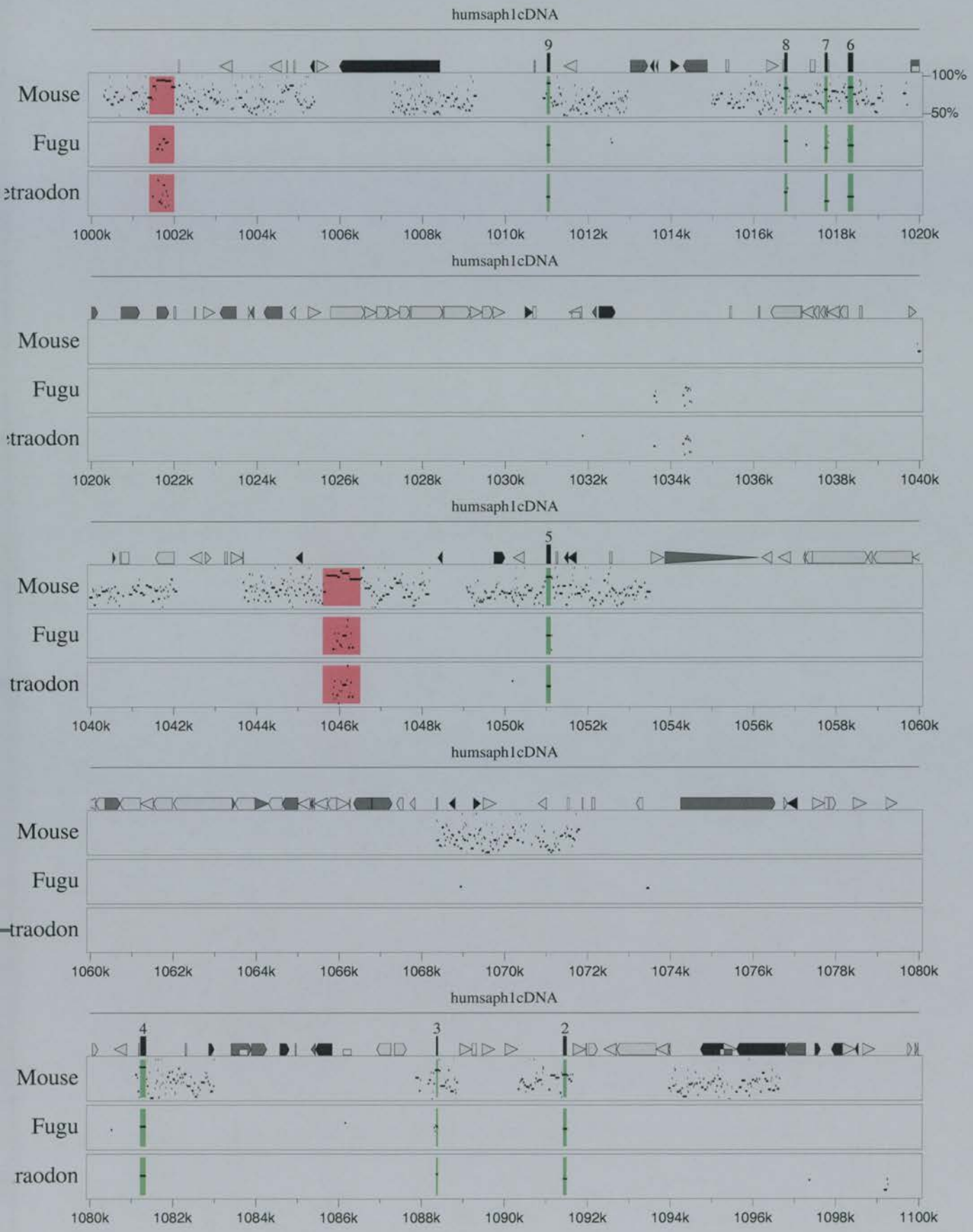

Mouse
Fugu
traodon
660k
662k
664k
666k
668k
670k
672k
674k
676k
678k
680k

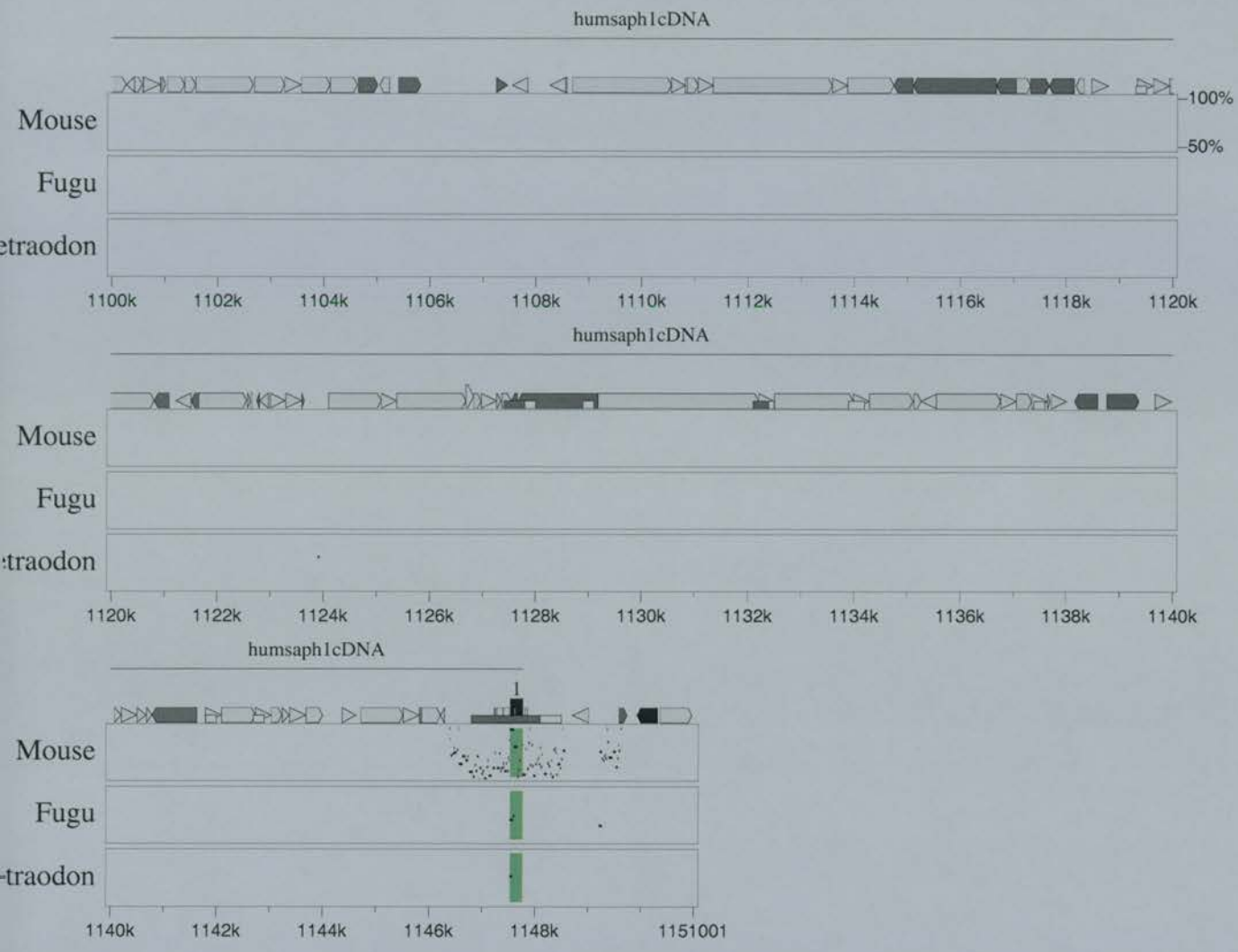

Mouse
Fugu
raodon
680k
682k
684k
686k
688k
690k
692k
694k
696k
698k
700k


Mouse
Fugu
etraodon
100%
50%
700k 702k 704k 706k 708k 710k 712k 714k 716k 718k 720k
Mouse
Fugu
traodon
720k 722k 724k 726k 728k 730k 732k 734k 736k 738k 740k
Mouse
Fugu
traodon
740k 742k 744k 746k 748k 750k 752k 754k 756k 758k 760k
Mouse
Fugu
traodon
760k 762k 764k 766k 768k 770k 772k 774k 776k 778k 780k
Mouse
Fugu
raodon
780k 782k 784k 786k 788k 790k 792k 794k 796k 798k 800k


Mouse
Fugu
etraodon
100%
50%
800k 802k 804k 806k 808k 810k 812k 814k 816k 818k 820k
Mouse
Fugu
traodon
820k 822k 824k 826k 828k 830k 832k 834k 836k 838k 840k
Mouse
Fugu
traodon
840k 842k 844k 846k 848k 850k 852k 854k 856k 858k 860k
Mouse
Fugu
traodon
860k 862k 864k 866k 868k 870k 872k 874k 876k 878k 880k
C7orf13mRNA
RNF32mRNA
1
1
2
3
Mouse
Fugu
raodon
880k 882k 884k 886k 888k 890k 892k 894k 896k 898k 900k

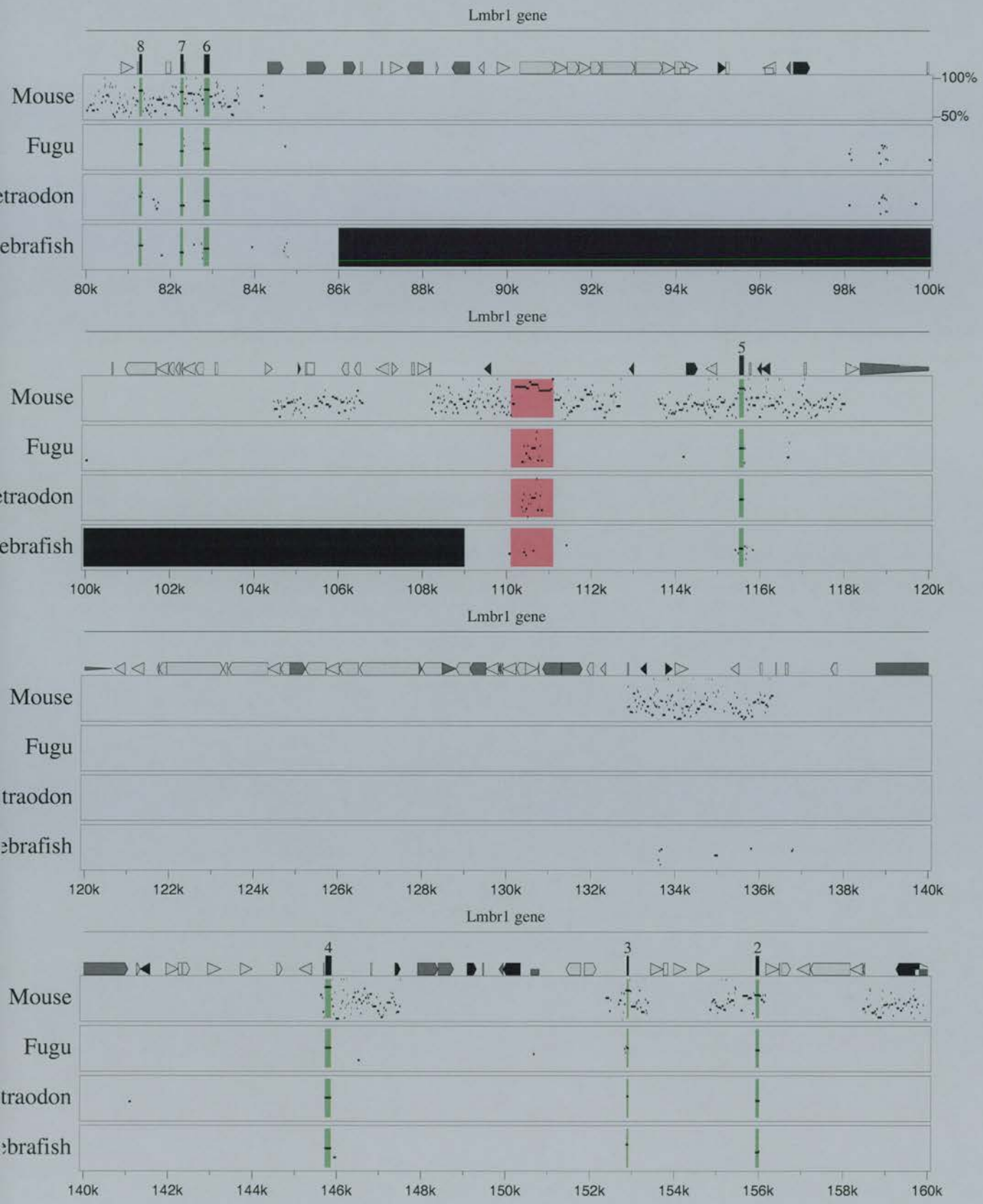

RNF32mRNA
4
5
6
7
Mouse
Fugu
etraodon
100%
50%
900k
902k
904k
906k
908k
910k
912k
914k
916k
918k
920k
RNF32mRNA
humsaph1cDNA
8
9
17
Mouse
Fugu
traodon
920k
922k
924k
926k
928k
930k
932k
934k
936k
938k
940k
humsaph1cDNA
16
Mouse
Fugu
traodon
940k
942k
944k
946k
948k
950k
952k
954k
956k
958k
960k
humsaph1cDNA
15
Mouse
Fugu
traodon
960k
962k
964k
966k
968k
970k
972k
974k
976k
978k
980k
humsaph1cDNA
13
12
11
10
Mouse
Fugu
raodon
980k
982k
984k
986k
988k
990k
992k
994k
996k
998k
1000k


humsaph1cDNA
9
8
7
6
Mouse
Fugu
etraodon
100%
50%
1000k
1002k
1004k
1006k
1008k
1010k
1012k
1014k
1016k
1018k
1020k
humsaph1cDNA
Mouse
Fugu
etraodon
1020k
1022k
1024k
1026k
1028k
1030k
1032k
1034k
1036k
1038k
1040k
humsaph1cDNA
5
Mouse
Fugu
traodon
1040k
1042k
1044k
1046k
1048k
1050k
1052k
1054k
1056k
1058k
1060k
humsaph1cDNA
Mouse
Fugu
traodon
1060k
1062k
1064k
1066k
1068k
1070k
1072k
1074k
1076k
1078k
1080k
humsaph1cDNA
4
3
2
Mouse
Fugu
raodon
1080k
1082k
1084k
1086k
1088k
1090k
1092k
1094k
1096k
1098k
1100k


humsaph1cDNA
Mouse
Fugu
etraodon
100%
50%
1100k
1102k
1104k
1106k
1108k
1110k
1112k
1114k
1116k
1118k
1120k
humsaph1cDNA
Mouse
Fugu
traodon
1120k
1122k
1124k
1126k
1128k
1130k
1132k
1134k
1136k
1138k
1140k
humsaph1cDNA
1
Mouse
Fugu
traodon
1140k
1142k
1144k
1146k
1148k
1151001

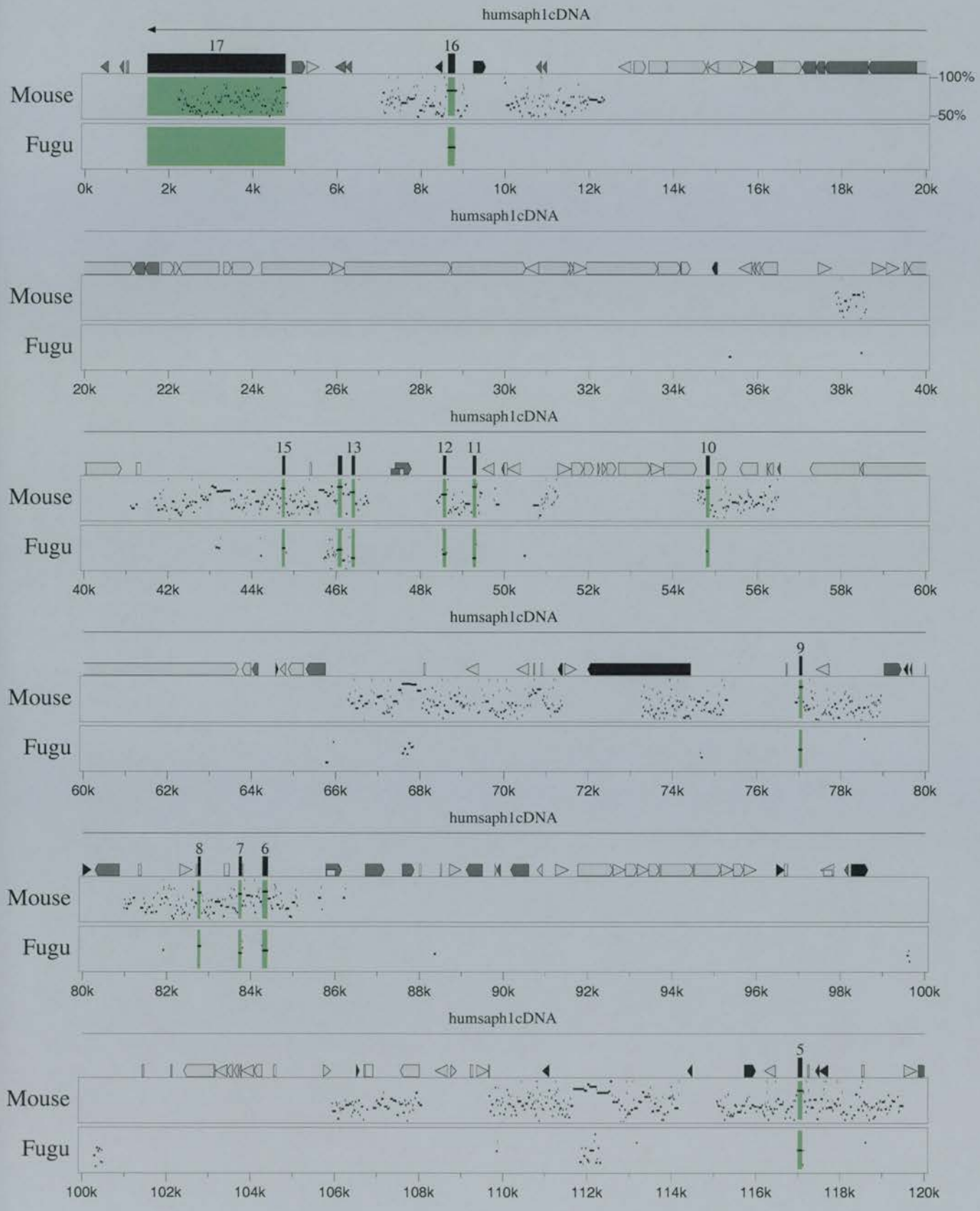
PIP plot *Lmbr1* only

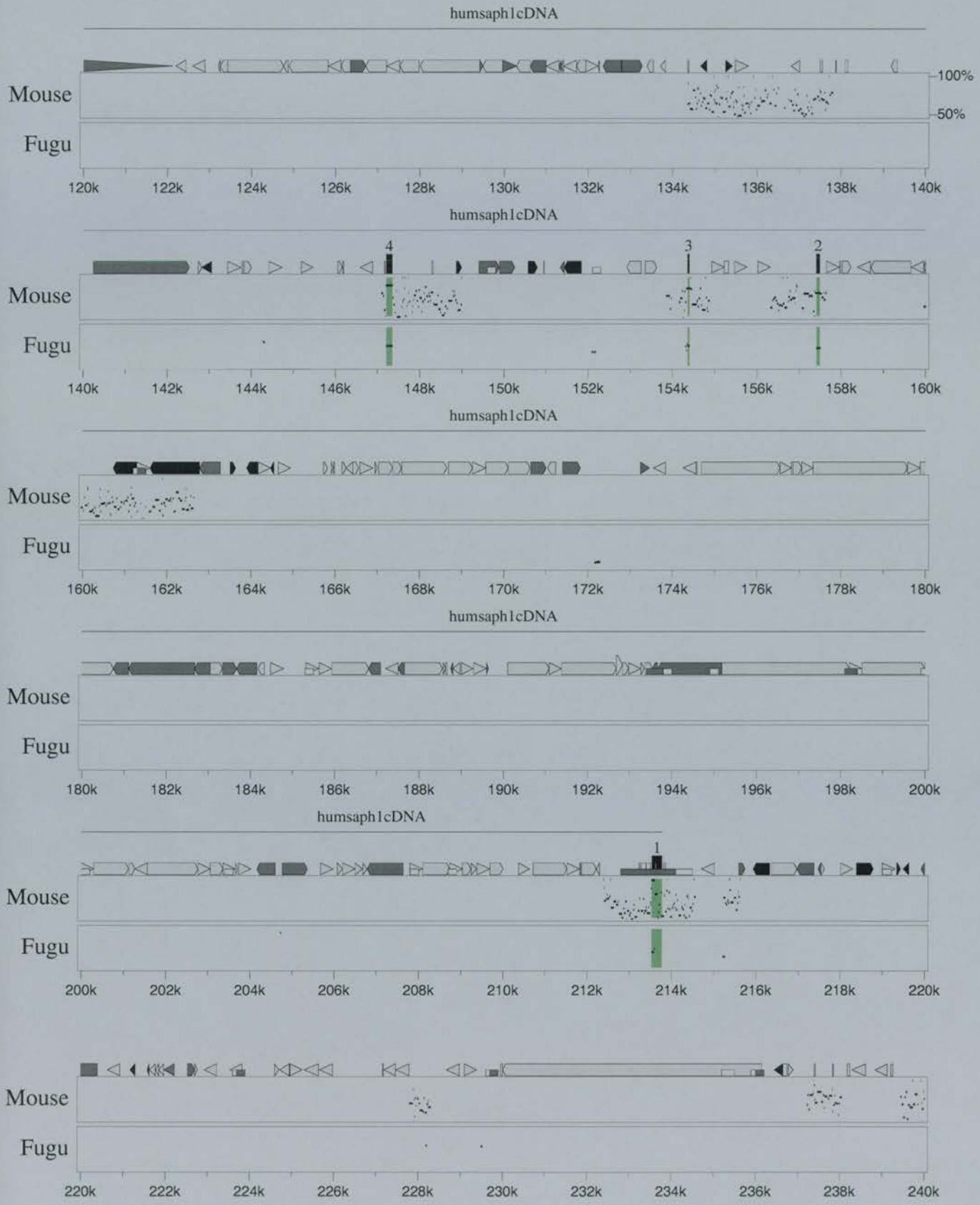

Lmbr1 gene
17
16
100%
50%
Mouse
Fugu
etraodon
ebrafish
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
Lmbr1 gene
Mouse
Fugu
traodon
ebrafish
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
Lmbr1 gene
15
13
12
11
10
Mouse
Fugu
traodon
ebrafish
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
Lmbr1 gene
9
Mouse
Fugu
traodon
ebrafish
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k

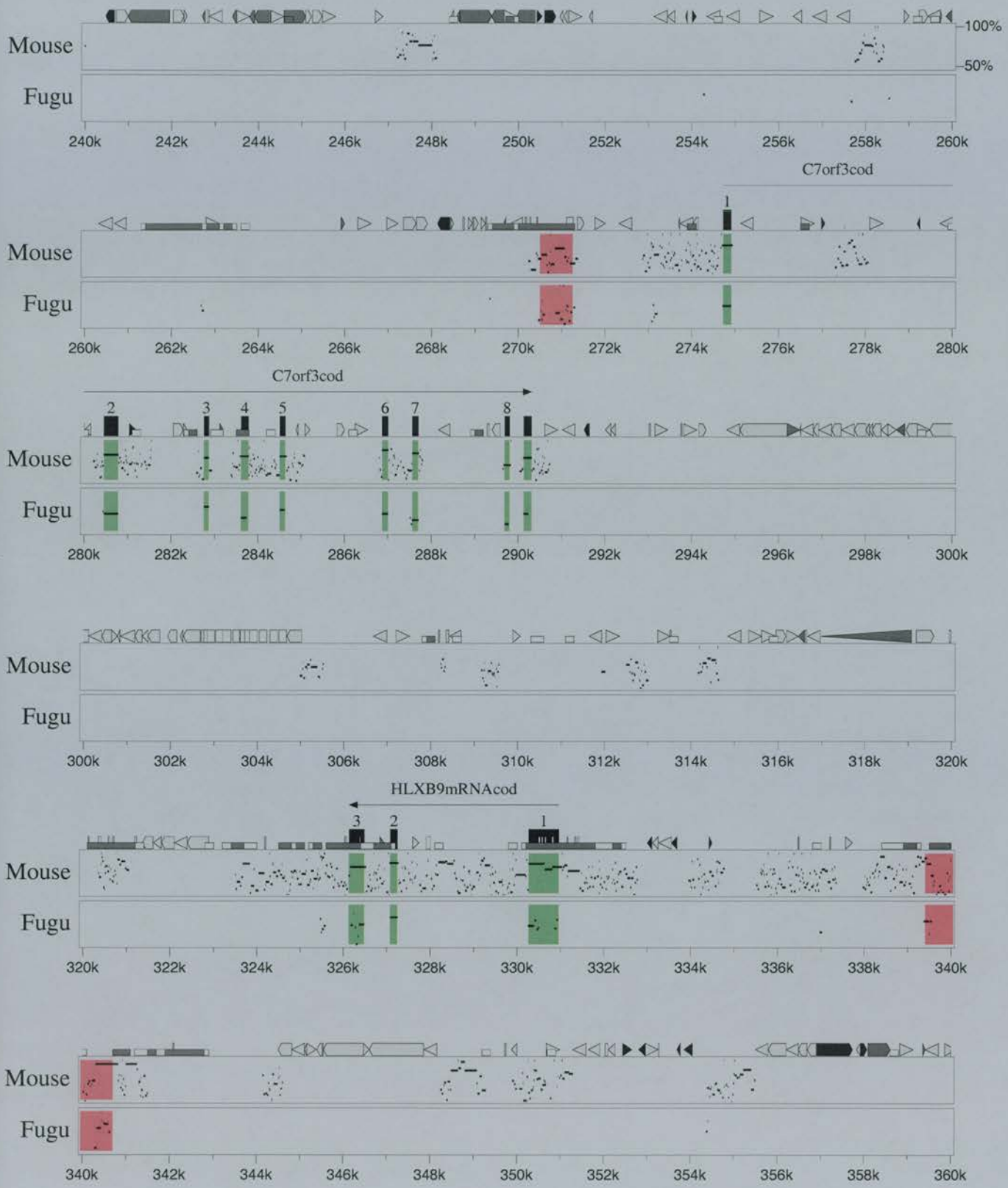

Lmbr1 gene
8
7
6
100%
50%
Mouse
Fugu
etraodon
ebrafish
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k
Lmbr1 gene
5
Mouse
Fugu
etraodon
ebrafish
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k
Lmbr1 gene
Mouse
Fugu
traodon
ebrafish
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k
Lmbr1 gene
4
3
2
Mouse
Fugu
traodon
ebrafish
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k

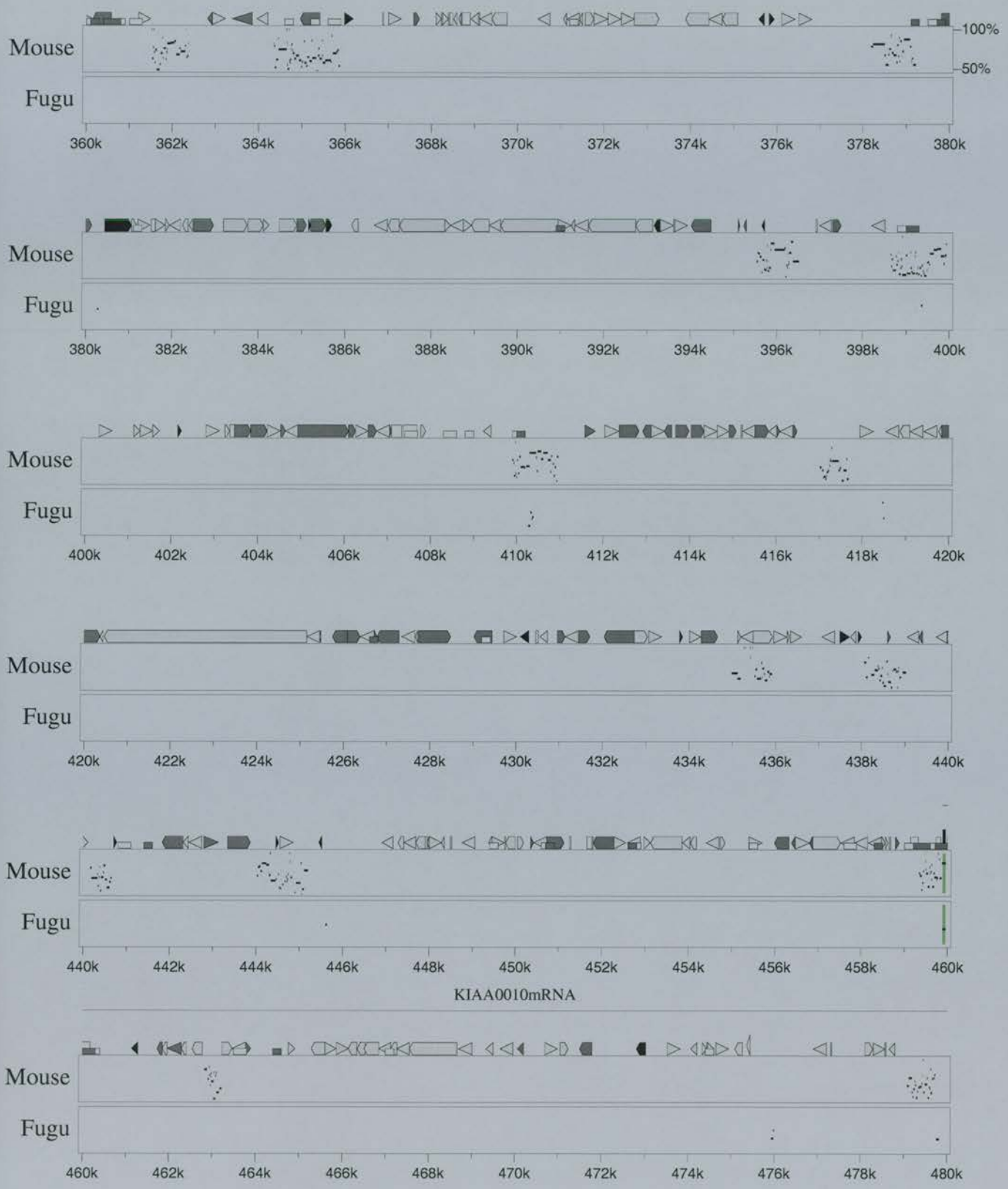

Lmbr1 gene
Mouse
Fugu
etraodon
ebrafish
100%
50%
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k
Lmbr1 gene
Mouse
Fugu
etraodon
ebrafish
180k
182k
184k
186k
188k
190k
192k
194k
196k
198k
200k
Lmbr1 gene
1
Mouse
Fugu
traodon
ebrafish
200k
202k
204k
206k
208k
210k
212291

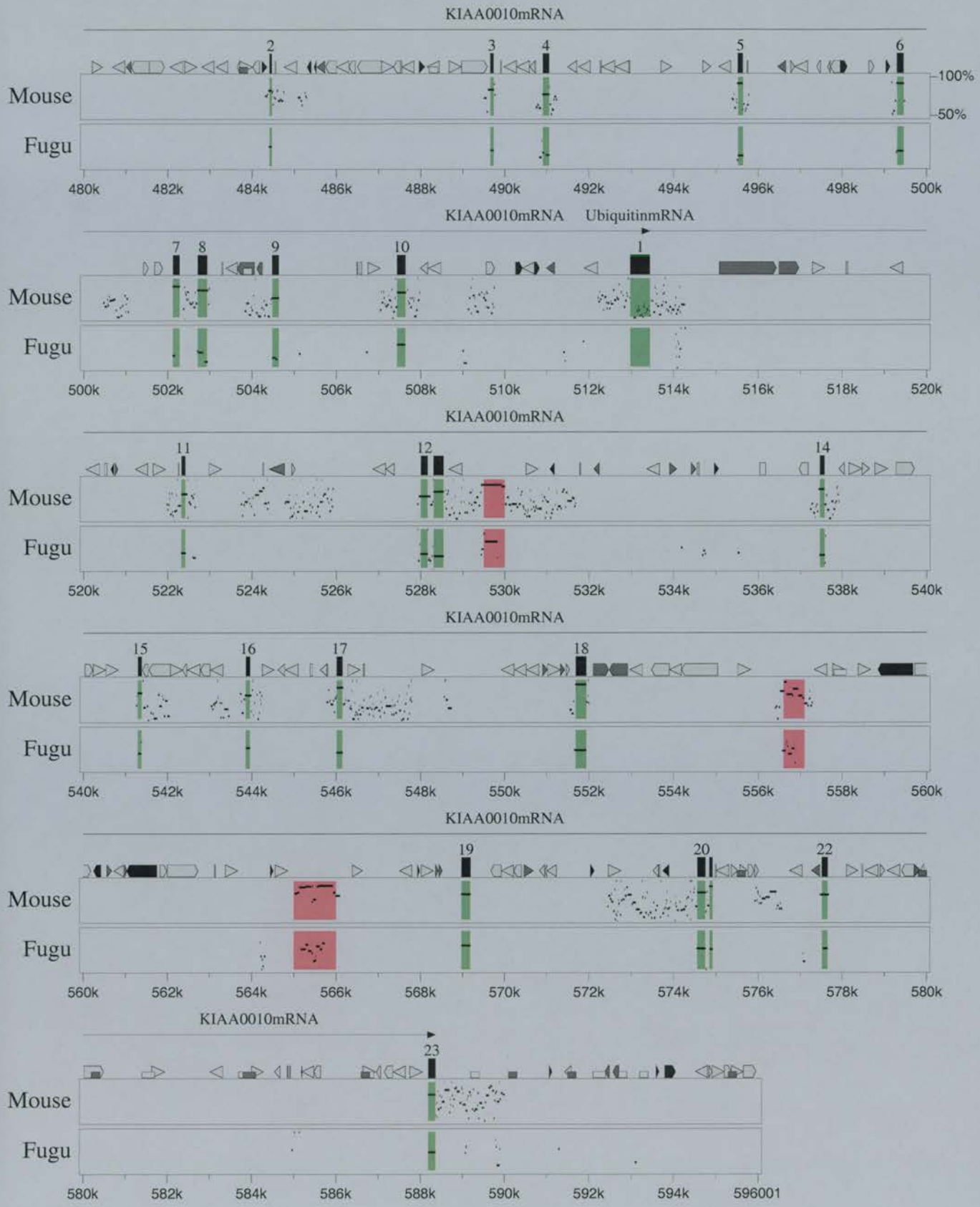
PIP plot *Lmbr1* to *KIAA0010*


humsaph1cDNA
17
16
Mouse
Fugu
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
humsaph1cDNA
Mouse
Fugu
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
humsaph1cDNA
15
13
12
11
10
Mouse
Fugu
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
humsaph1cDNA
9
Mouse
Fugu
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
humsaph1cDNA
8
7
6
Mouse
Fugu
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k
humsaph1cDNA
5
Mouse
Fugu
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k


humsaph1cDNA
Mouse
Fugu
100%
50%
120k 122k 124k 126k 128k 130k 132k 134k 136k 138k 140k
humsaph1cDNA
4
3
2
Mouse
Fugu
140k 142k 144k 146k 148k 150k 152k 154k 156k 158k 160k
humsaph1cDNA
Mouse
Fugu
160k 162k 164k 166k 168k 170k 172k 174k 176k 178k 180k
humsaph1cDNA
Mouse
Fugu
180k 182k 184k 186k 188k 190k 192k 194k 196k 198k 200k
humsaph1cDNA
1
Mouse
Fugu
200k 202k 204k 206k 208k 210k 212k 214k 216k 218k 220k
Mouse
Fugu
220k 222k 224k 226k 228k 230k 232k 234k 236k 238k 240k

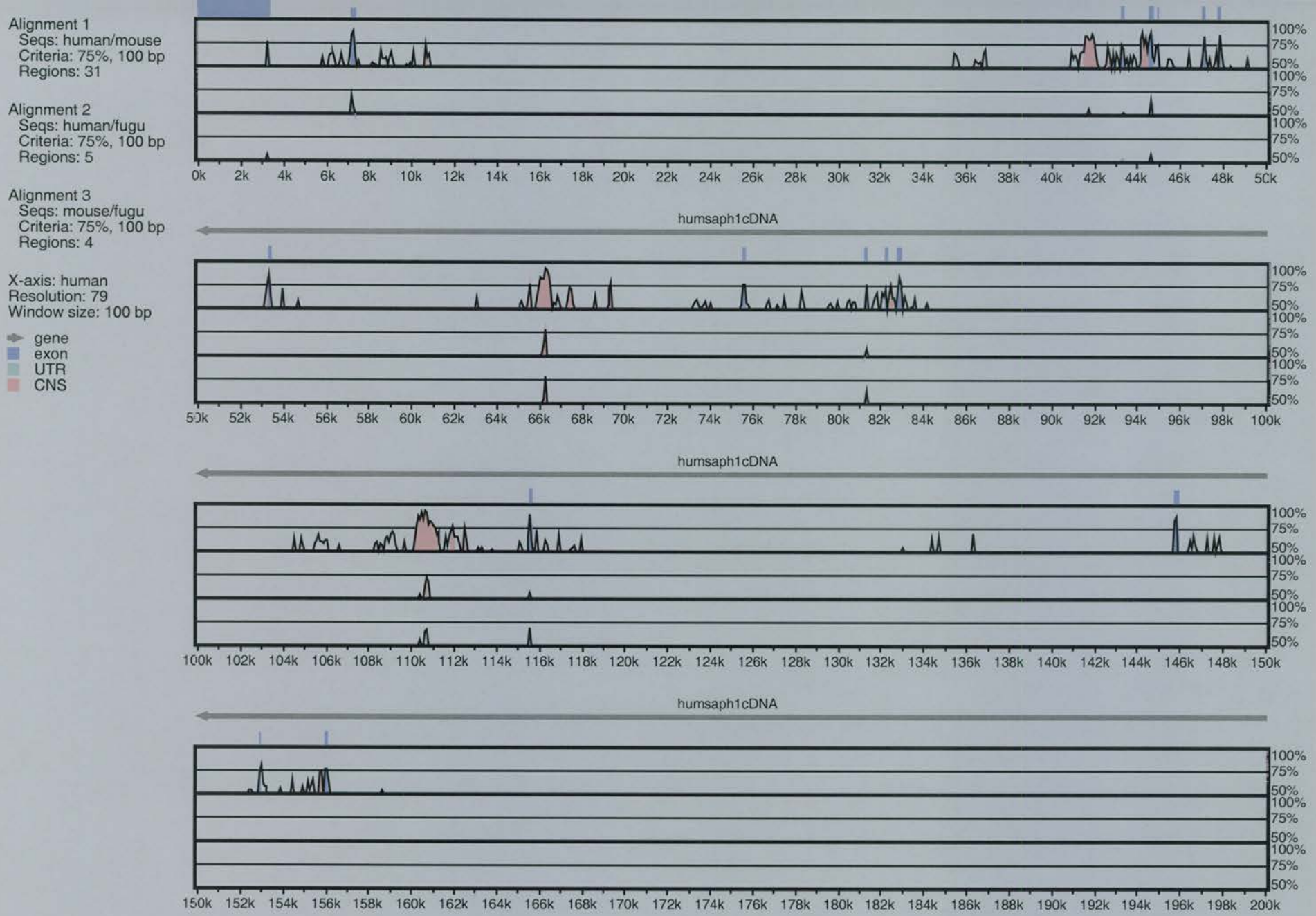

Mouse
Fugu
100%
50%
240k 242k 244k 246k 248k 250k 252k 254k 256k 258k 260k
C7orf3cod
1
260k 262k 264k 266k 268k 270k 272k 274k 276k 278k 280k
C7orf3cod
2 3 4 5 6 7 8
280k 282k 284k 286k 288k 290k 292k 294k 296k 298k 300k
300k 302k 304k 306k 308k 310k 312k 314k 316k 318k 320k
HLXB9mRNAcod
3 2 1
320k 322k 324k 326k 328k 330k 332k 334k 336k 338k 340k
340k 342k 344k 346k 348k 350k 352k 354k 356k 358k 360k

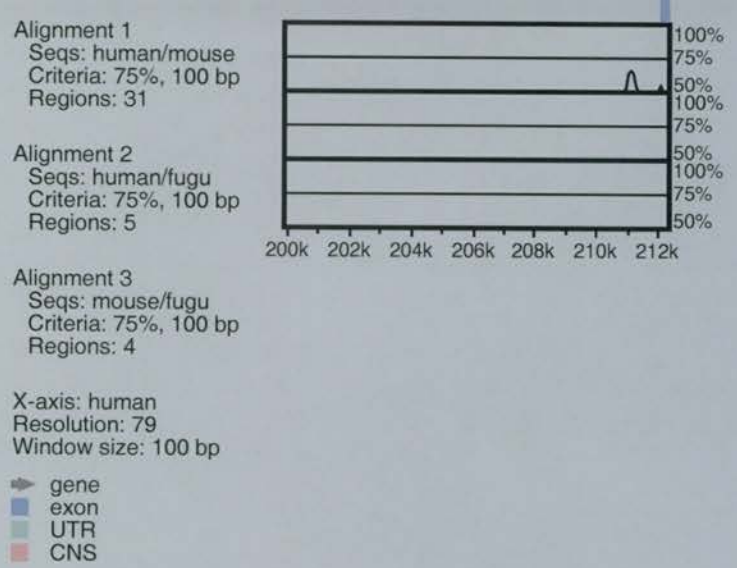

Mouse
Fugu
100%
50%
360k 362k 364k 366k 368k 370k 372k 374k 376k 378k 380k
Mouse
Fugu
380k 382k 384k 386k 388k 390k 392k 394k 396k 398k 400k
Mouse
Fugu
400k 402k 404k 406k 408k 410k 412k 414k 416k 418k 420k
Mouse
Fugu
420k 422k 424k 426k 428k 430k 432k 434k 436k 438k 440k
Mouse
Fugu
440k 442k 444k 446k 448k 450k 452k 454k 456k 458k 460k
KIAA0010mRNA
Mouse
Fugu
460k 462k 464k 466k 468k 470k 472k 474k 476k 478k 480k


KIAA0010mRNA
2
3
4
5
6
Mouse
Fugu
100%
50%
480k
482k
484k
486k
488k
490k
492k
494k
496k
498k
500k
KIAA0010mRNA
UbiquitinmRNA
7
8
9
10
1
Mouse
Fugu
500k
502k
504k
506k
508k
510k
512k
514k
516k
518k
520k
KIAA0010mRNA
11
12
14
Mouse
Fugu
520k
522k
524k
526k
528k
530k
532k
534k
536k
538k
540k
KIAA0010mRNA
15
16
17
18
Mouse
Fugu
540k
542k
544k
546k
548k
550k
552k
554k
556k
558k
560k
KIAA0010mRNA
19
20
22
Mouse
Fugu
560k
562k
564k
566k
568k
570k
572k
574k
576k
578k
580k
KIAA0010mRNA
23
Mouse
Fugu
580k
582k
584k
586k
588k
590k
592k
594k
596001

Appendix 2

VISTA plot of genomic region 2


Alignment 1
Seqs: human/mouse
Criteria: 75%, 100 bp
Regions: 31
Alignment 2
Seqs: human/fugu
Criteria: 75%, 100 bp
Regions: 5
Alignment 3
Seqs: mouse/fugu
Criteria: 75%, 100 bp
Regions: 4
X-axis: human
Resolution: 79
Window size: 100 bp
gene
exon
UTR
CNS
100%
75%
50%
0k 2k 4k 6k 8k 10k 12k 14k 16k 18k 20k 22k 24k 26k 28k 30k 32k 34k 36k 38k 40k 42k 44k 46k 48k 50k
humsaph1cDNA
50k 52k 54k 56k 58k 60k 62k 64k 66k 68k 70k 72k 74k 76k 78k 80k 82k 84k 86k 88k 90k 92k 94k 96k 98k 100k
humsaph1cDNA
100k 102k 104k 106k 108k 110k 112k 114k 116k 118k 120k 122k 124k 126k 128k 130k 132k 134k 136k 138k 140k 142k 144k 146k 148k 150k
humsaph1cDNA
150k 152k 154k 156k 158k 160k 162k 164k 166k 168k 170k 172k 174k 176k 178k 180k 182k 184k 186k 188k 190k 192k 194k 196k 198k 200k


Alignment 1
Seqs: human/mouse
Criteria: 75%, 100 bp
Regions: 31
Alignment 2
Seqs: human/fugu
Criteria: 75%, 100 bp
Regions: 5
Alignment 3
Seqs: mouse/fugu
Criteria: 75%, 100 bp
Regions: 4
X-axis: human
Resolution: 79
Window size: 100 bp
gene
exon
UTR
CNS
100%
75%
50%
100%
75%
50%
100%
75%
50%
200k
202k
204k
206k
208k
210k
212k